(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 September 2003 (18.09.2003)

PCT

(10) International Publication Number
WO 03/076658 A2

(51) International Patent Classification[7]: C12Q 1/68, C12N 15/00, 5/00, G01N 33/50, A61P 25/16

(21) International Application Number: PCT/IB02/04276

(22) International Filing Date: 14 October 2002 (14.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/363,220 8 March 2002 (08.03.2002) US

(71) Applicant *(for all designated States except US)*: **DECODE GENETICS EHF.** [IS/IS]; Sturlugotu 8, IS-101 Reykjavik (IS).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **HICKS, Andrew, A.** [GB/GB]; 104 Bankton Park East, Livingston EH54 9BN (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/076658 A2

(54) Title: A SUSCEPTIBILITY GENE FOR LATE-ONSET IDIOPATHIC PARKINSON'S DISEASE

(57) Abstract: Methods of diagnosis of susceptibility to Parkinson's disease; assays for agents that alter the activity of a Parkinson's disease polypeptide, (e.g., PARK8 polypeptide) or which identify PARK8 binding agents, and the agents or binding agents identified by the assays; PARK8 therapeutic agents, including the PARK8 nucleic acids, PARK8 polypeptides, or agents that alter the activity of an PARK8 polypeptides; pharmaceutical compositions comprising the PARK8 therapeutic agents; as well as methods of therapy of Parkinson' s disease are disclosed.

A SUSCEPTIBILITY GENE FOR LATE-ONSET IDIOPATHIC PARKINSON'S DISEASE

RELATED APPLICATION(S)

This application claims the benefit of U.S. Provisional Application No. 60/363,220, filed March 8, 2002. The entire teachings of the above application are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Parkinson's disease is a neurodegenerative disorder of unclear etiology and pathogenesis that usually occurs in middle-age or later with an onset of symptoms typically after 60 years of age. The signs of the disease include bradykinesia, muscular rigidity, resting tremor, and postural instability. In some rare families characterized by early-onset and a Mendelian pattern of transmission, the role of genetic factors has been clearly established over recent years. In families with autosomal dominant early-onset Parkinson's disease, the first gene to be identified was α-synuclein, (Polymeropoulos, M.H., *et al.*, *Science.* 276: 2045-7 (1997)). Two additional loci have been reported in such families, including *PARK3* on chromosome 2p13 (Gasser, T., *et al.*, *Nat. Genet.* 18:262-5 (1998)), and *PARK4* on chromosome 4p14-16.3, (Farrer, M., *et al.*, *Hum. Mol. Genet.* 8:81-5 (1999)), and a mutation in the ubiquitin carboxy-terminal hydrolase L1 gene in one German family have been reported (Leroy E., *et al.*, *Nature.* 395: 451-2 (1998)). Other families studied report no linkage to these regions (Farrer, M., *et al.*, *Mov. Disord.* 15:1075-83 (2000)). In addition, the Parkin gene on chromosome 6q25.2-27 *(PARK2)* has been shown to be involved in juvenile parkinsonism (Matsumine, H., *et al.*, *Am. J. Hum. Genet.*, 60:588-96 (1997)), and in families demonstrating an autosomal recessive mode of inheritance as well as some sporadic early-onset cases (Kitada T., *et al.*, *Nature.* 392:605-8 (1998), Abbas, N., *et al.*, *Hum. Mol. Genet.* 8:567-74 (1999), Tassin, J., *et al.*, *Am. J. Hum. Genet.*, 63:88-94 (1998) and Lucking, C.B., *et al.*, *N. Engl. J. Med.* 342:1560-7 (2000)). Additionally, two loci on chromosome

1p35-p36, *PARK6 and PARK7,* have been recently reported to contain another susceptibility gene in families with early-onset recessive Parkinson's disease, (Valente, E.M., *et al., Am. J. Hum. Genet.* 68:895-900 (2001), van Dujin, C.M., *et al., Am. J. Hum. Genet.* 69:629-634 (2001)).

The role of genetic factors in both early onset, autosomal dominant and recessive Parkinson's disease have been clearly established, but, these forms represent only a minor part of the total disease spectrum. The common late-onset form of Parkinson's disease does not follow a Mendelian pattern of inheritance and some twin studies have spawned the conclusion that there are no genetic factors involved in its risk, (Tanner, C.M., *et al., JAMA.* 281:341-6 (1999)). The disease is a complex, multifactorial disease resulting from interaction between one or more genes and the environment.

Elucidation of the relative role of genetic and environmental factors is needed. In addition to family information for the study of Parkinson's disease, it is desirable to develop diagnostic methods for the early diagnosis of the disease or to establish predisposition for the development of Parkinson's disease. Understanding the role of genetic and environmental factors can lead to improved therapeutics.

SUMMARY OF THE INVENTION

As described herein, a region on chromosome 1p32 designated as *PARK8* comprising the gene that encodes ring finger protein 11 (hereinafter referred to as "RNF 11") has been correlated through human linkage studies to Parkinson's disease. The present invention relates the use of nucleic acid molecules comprising the *PARK8* gene and encoded proteins for diagnostic and therapeutic methods.

The invention also relates to methods of diagnosing a susceptibility to Parkinson's disease in an individual, comprising detecting a polymorphism in the *PARK8* gene, wherein the presence of the polymorphism in the gene is indicative of a susceptibility to Parkinson's disease. The invention also pertains to methods of detecting an alteration in the expression or composition of a polypeptide encoded by *PARK8* gene in a test sample, in comparison with the expression or composition of a polypeptide encoded by *PARK8* gene in a control sample, wherein the presence of an

alteration in expression or composition of the polypeptide in the test sample is indicative of a susceptibility to Parkinson's disease.

This invention further pertains to methods of identifying an agent which alters activity of a polypeptide encoded by a Parkinson's disease gene, comprising contacting the polypeptide or a derivative or fragment thereof, with an agent to be tested; assessing the level of activity of the polypeptide or derivative or fragment thereof; and comparing the level of activity with a level of activity of the polypeptide or active derivative or fragment thereof in the absence of the agent, wherein if the level of activity of the polypeptide or derivative or fragment thereof in the presence of the agent differs, by an amount that is statistically significant, from the level in the absence of the agent, then the agent is an agent that alters activity of the polypeptide are described. Agents which alter activity of a polypeptide encoded by a Parkinson's disease gene, identified according to this method are also described. The agents which alter activity of a polypeptide encoded by a Parkinson's disease gene, can be: a Parkinson's disease receptor; a Parkinson's disease binding agent; a peptidomimetic; a fusion protein; a prodrug; an antibody; and a ribozyme. These agents can also be utilized in a method of altering activity of a polypeptide encoded by a Parkinson's disease gene.

The invention also relates to a method of identifying an agent which alters interaction of the polypeptide encoded by a Parkinson's disease gene with a Parkinson's disease binding agent, comprising contacting the polypeptide or a derivative or fragment thereof, and the binding agent, with an agent to be tested; assessing the interaction of the polypeptide or derivative or fragment thereof with the binding agent; and comparing the level of interaction with a level of interaction of the polypeptide or derivative or fragment thereof with the binding agent in the absence of the agent, wherein if the level of interaction of the polypeptide or derivative or fragment thereof in the presence of the agent differs, by an amount that is statistically significant, from the level of interaction in the absence of the agent, then the agent is an agent that alters interaction of the polypeptide with the binding agent. Agents which alter interaction of a Parkinson's disease polypeptide with a Parkinson's disease binding agent are also described and can be a Parkinson's

disease receptor; a second Parkinson's disease binding agent; a peptidomimetic; a fusion protein; a prodrug; an antibody; and a ribozyme.

A method of altering interaction of a Parkinson's disease polypeptide with a Parkinson's disease binding agent, comprising contacting the Parkinson's disease polypeptide and/or the Parkinson's disease binding agent with an agent are also described.

The invention additionally relates to a method of identifying an agent which alters expression of a Parkinson's disease gene, comprising contacting a solution containing a nucleic acid of the invention or a derivative or fragment thereof, with an agent to be tested; assessing the level of expression of the nucleic acid, derivative or fragment; and comparing the level of expression with a level of expression of the nucleic acid, derivative or fragment in the absence of the agent, wherein if the level of expression of the nucleotide, derivative or fragment in the presence of the agent differs, by an amount that is statistically significant, from the expression in the absence of the agent, then the agent is an agent that alters expression of a Parkinson's disease gene, as well as agents identified or identifiable by this method.

The invention also relates to a method of identifying an agent which alters expression of a Parkinson's disease gene, comprising contacting a solution containing a nucleic acid comprising the promoter region of a Parkinson's disease gene operably linked to a reporter gene, with an agent to be tested; assessing the level of expression of the reporter gene; and comparing the level of expression with a level of expression of the reporter gene in the absence of the agent, wherein if the level of expression of the reporter gene in the presence of the agent differs, by an amount that is statistically significant, from the level of expression in the absence of the agent, then the agent is an agent that alters expression of a Parkinson's disease gene. The invention also relates to agents identifiable by this method.

The invention further relates to a method of identifying an agent which alters expression of a Parkinson's disease gene, comprising contacting a solution containing a nucleic acid of the invention or a derivative or fragment thereof with an agent to be tested; assessing expression of the nucleic acid, derivative or fragment; and

comparing expression with expression of the nucleic acid, derivative or fragment in the absence of the agent, wherein if expression of the nucleotide, derivative or fragment in the presence of the agent differs, by an amount that is statistically significant, from the expression in the absence of the agent, then the agent is an agent that alters expression of a Parkinson's disease gene. The expression of the nucleotide, derivative or fragment in the presence of the agent comprises expression of one or more variant(s) that differ in kind or in quantity from the expression of variant(s) in the absence of the agent. Agents identified by such a method are also contemplated.

In another embodiment of the invention, a method of identifying a polypeptide which interacts with a Parkinson's disease polypeptide, comprising for example, employing a yeast two hybrid system using a first vector which comprises a nucleic acid encoding a DNA binding domain and a Parkinson's disease polypeptide, variant, or fragment or derivative thereof, and a second vector which comprises a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a test polypeptide, wherein if transcriptional activation occurs in the yeast two hybrid system, the test polypeptide is a polypeptide which interacts with a Parkinson's disease polypeptide.

Additionally, the invention pertains to pharmaceutical compositions comprising the nucleic acids of the invention, the polypeptides of the invention, and/or the agents that alter activity of the *PARK8* polypeptides. The invention further pertains to methods of treating Parkinson's disease in an individual, comprising administering a Parkinson's disease therapeutic agent to the individual, in a therapeutically effective amount.

BRIEF DESCRIPTION OF THE FIGURES

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings.

FIG. 1 shows a family pedigree affected by Parkinson's disease.

FIG. 2 depicts multipoint allele lod score of chromosome 1 with versus Z_{lr}

score, with extra microsatellite markers within *PARK8*.

FIGS. 3A-3L depict the nucleic acid sequence of human RNF 11 (SEQ ID NO:1) and the amino acid sequence of the encoded protein RNF 11 (SEQ ID NO: 2).

DETAILED DESCRIPTION OF THE INVENTION

Extensive genealogical information for a population with population-based lists of patients have been combined with powerful genome sharing methods to map a major locus in common late-onset Parkinson's disease.

A genome-wide scan on patients, related within 6 meiotic events (6 meiosis separate second cousins), diagnosed within Parkinson's disease and their unaffected relatives has been completed. This linkage investigation revealed a region on chromosome 1p32 designated *PARK8*, to be associated with Parkinson's disease. This locus is distinct from previously published loci contributing to familial Parkinson's disease. *PARK6* and *PARK7* are also located on chromosome 1p but appear to be at least 35 centimorgans away from *PARK8* (Valente, E.M., *et al.*, *Am. J. Hum. Genet.* 68:895-900 (2001); van Dujin, C.M., *et al.*, *Am. J. Hum. Genet.* 69:629-634 (2001)). From the physical map generated herein and public data, there appear to be at least 30 known or possible genes in the region of *PARK8*. This locus does not correspond to known susceptibility loci for Parkinson's disease, and the work described herein represents the first mapping of a gene for common Parkinson's disease. Identification of the susceptibility genes in the region can pave the way for a better understanding of the disease process, which in turn can lead to improved diagnosis and therapeutics.

Described herein is the first known linkage study of Parkinson's disease showing a connection to chromosome 1p32. Based on the linkage studies conducted, a direct relationship between the *PARK8* locus, specifically the RNF 11 gene, and Parkinson's disease has been discovered.

NUCLEIC ACIDS OF THE INVENTION

Genomic DNA from patient samples were sequences in the region of the *PARK8* locus and the gene encoding ring finger protein 11 was identified. The

nucleic acid sequence is shown in SEQ ID NO: 1 and FIGS. 3A to 3L. Seven SNP's in the promoter region of the gene and one SNP next to the ATG start codon have been identified. Based upon this discovery, the invention pertains to an isolated nucleic acid molecule comprising *PARK8* fragment or variant thereof and further comprising at least one single nucleotide polymorphism selected from the group consisting of G at nucleotide 1196, C at nucleotide 1393, T at nucleotide 2064, T at nucleotide 2218, C at nucleotide 2435, C at nucleotide 2599, G at nucleotide 2703 and G at nucleotide 3181. The nucleotide numbering is relative to SEQ ID NO: 1.

The isolated nucleic acid molecules of the present invention can be RNA, for example, mRNA, or DNA, such as cDNA and genomic DNA. DNA molecules can be double-stranded or single-stranded; single-stranded RNA or DNA can be either the coding, or sense, strand or the non-coding, or antisense, strand. The nucleic acid molecule can include all or a portion of the coding sequence of the gene and can further comprise additional non-coding sequences such as introns and non-coding 3' and 5' sequences (including regulatory sequences, for example). Additionally, the nucleic acid molecule can be fused to a marker sequence, for example, a sequence that encodes a polypeptide to assist in isolation or purification of the polypeptide. Such sequences include, but are not limited to, those which encode a glutathione-S-transferase (GST) fusion protein and those which encode a hemagglutinin A (HA) polypeptide marker from influenza.

An "isolated" nucleic acid molecule, as used herein, is one that is separated from nucleic acids which normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (*e.g.*, as in an RNA library). For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity,

for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid molecule comprises at least about 50, 80, 90 or 95% (on a molar basis) of all macromolecular species present. With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotides which flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence which is synthesized chemically or by recombinant means. Therefore, recombinant DNA contained in a vector are included in the definition of "isolated" as used herein. Also, isolated nucleotide sequences include recombinant DNA molecules in heterologous organisms, as well as partially or substantially purified DNA molecules in solution. *In vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention are also encompassed by "isolated" nucleotide sequences. Such isolated nucleotide sequences are useful in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (*e.g.*, from other mammalian species), for gene mapping (*e.g.*, by *in situ* hybridization with chromosomes), or for detecting expression of the gene in tissue (*e.g.*, human tissue), such as by Northern blot analysis.

The present invention also pertains to variant nucleic acid molecules which are not necessarily found in nature but which encode an RNF 11 polypeptide (*e.g.*, a polypeptide having the amino acid sequence of any one of Parkinson's disease genes

or another variant of a RNF 11 polypeptide). Thus, for example, DNA molecules which comprise a sequence that is different from the naturally-occurring nucleotide sequence but which, due to the degeneracy of the genetic code, encode an RNF 11 polypeptide of the present invention are also the subject of this invention. The invention also encompasses nucleotide sequences encoding portions (fragments), or encoding variant polypeptides such as analogues or derivatives of a RNF 11 polypeptide. Such variants can be naturally-occurring, such as in the case of allelic variation or single nucleotide polymorphisms, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion and substitution of one or more nucleotides which can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably the nucleotide (and/or resultant amino acid) changes are silent or conserved; that is, they do not alter the characteristics or activity of the RNF 11 polypeptide. In one preferred embodiment, the nucleotide sequences are fragments that comprise one or more polymorphic microsatellite markers.

Other alterations of nucleic acid molecules of the invention can include, for example, labelling, methylation, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids). Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequences via hydrogen bonding and other chemical interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also pertains to nucleic acid molecules which hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (*e.g.*, nucleic acid molecules which specifically hybridize to a nucleotide sequence encoding polypeptides described herein, and, optionally, have an activity of the polypeptide). In one embodiment, the invention

includes variants described herein which hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleotide sequence comprising a nucleotide sequence selected from SEQ ID NO: 1 or the complement of SEQ ID NO: 1. In another embodiment, the invention includes variants described herein which hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleotide sequence encoding an amino acid sequence encoded by SEQ ID NO: 1. In a preferred embodiment, the variant which hybridizes under high stringency hybridizations has an activity of RNF 11 (*e.g.*, binding activity).

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (*e.g.*, under high stringency conditions). "Specific hybridization," as used herein, refers to the ability of a first nucleic acid to hybridize to a second nucleic acid in a manner such that the first nucleic acid does not hybridize to any nucleic acid other than to the second nucleic acid (*e.g.*, when the first nucleic acid has a higher similarity to the second nucleic acid than to any other nucleic acid in a sample wherein the hybridization is to be performed). "Stringency conditions" for hybridization is a term of art which refers to the incubation and wash conditions, *e.g.*, conditions of temperature and buffer concentration, which permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be perfectly (*i.e.*, 100%) complementary to the second, or the first and second may share some degree of complementarity which is less than perfect (*e.g.*, 70%, 75%, 85%, 95%). For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity. "High stringency conditions", "moderate stringency conditions" and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6.3.6 in "*Current Protocols in Molecular Biology*" (Ausubel, F.M. *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (2001), the entire teachings of which are incorporated by reference herein). The exact conditions which determine the stringency of hybridization depend not only on ionic strength (*e.g.*, 0.2XSSC, 0.1XSSC), temperature (*e.g.*, room temperature, 42°C, 68°F) and the concentration of destabilizing agents such as

formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules. Typically, conditions are used such that sequences at least about 60%, at least about 70%, at least about 80%, at least about 90% or at least about 95% or more identical to each other remain hybridized to one another. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (*e.g.*, selectively) with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and S.A. Aaronson, *Methods in Enzymology*, 200:546-556 (1991). Also, in, Ausubel, *et al., Current Protocols in Molecular Biology*, John Wiley & Sons, (2001), which describes the determination of washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in T_m of ~17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate or low stringency, depending on the level of mismatch sought.

For example, a low stringency wash can comprise washing in a solution containing 0.2XSSC/0.1% SDS for 10 min at room temperature; a moderate stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2XSSC/0.1% SDS for 15 min at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1XSSC/0.1%SDS for 15 min at 68°C. Furthermore, washes can be performed repeatedly or sequentially to

obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 60%, and even more preferably at least 70%, 80% or 90% of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) as described in Altschul *et al.*, *Nucleic Acids Res.*, 25:389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at score=100, wordlength=12, or can be varied (*e.g.*, W=5 or W=20).

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti, *Comput. Appl. Biosci.*,

10:3-5 (1994); and FASTA described in Pearson and Lipman, *PNAS,* 85:2444-8 (1988).

In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG-software package (available at http://www.accelrys.com/about/gcg.html) using either a Blossom 63 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. In yet another embodiment, the percent identity between two nucleic acid sequences can be accomplished using the GAP program in the GCG-software package (available at http://www.accelrys.com/about/gcg.html), using a gap weight of 50 and a length weight of 3.

The present invention also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence comprising a nucleotide sequence of SEQ ID NO: 1 and the complement of SEQ ID NO: 1, and also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence encoding an amino acid sequence of SEQ ID NO: 2. The nucleic acid fragments of the invention are at least about 15, preferably at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic polypeptides described herein are particularly useful, such as for the generation of antibodies as described below.

In a related aspect, the nucleic acid fragments of the invention are used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid molecules. Such probes and primers include polypeptide nucleic acids, as described in Nielsen *et al.*, *Science*, 254:1497-1500 (1991). As also used herein, the term "primer" in particular refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (*e.g.*, PCR, LCR) including, but not limited to those described herein.

Typically, a probe or primer comprises a region of nucleotide sequence that hybridizes to at least about 15, typically about 20-25, and more typically about 40, 50 or 75, consecutive nucleotides of a nucleic acid molecule comprising a contiguous nucleotide sequence selected from: SEQ ID NO: 1, the complement of SEQ ID NO: 1, or a sequence encoding an amino acid sequence of SEQ ID NO: 2 In preferred embodiments, a probe or primer comprises 100 or fewer nucleotides, preferably from 6 to 50 nucleotides, preferably from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence, preferably at least 80% identical, more preferably at least 90% identical, even more preferably at least 95% identical, or even capable of selectively hybridizing to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. Often, the probe or primer further comprises a label, *e.g.*, radioisotope, fluorescent compound, enzyme or enzyme co-factor.

The nucleic acid molecules of the invention such as those described above can be identified and isolated using standard molecular biology techniques and the sequence information provided in SEQ ID NO: 1, and/or SEQ ID NO: 2. For example, nucleic acid molecules can be amplified and isolated by the polymerase chain reaction using synthetic oligonucleotide primers designed based on one or more of the sequences provided in SEQ ID NO: 1 and/or the complement of SEQ ID NO: 1, or designed based on nucleotides based on sequences encoding one or more of the amino acid sequences provided in SEQ ID NO: 2. See generally *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (Eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.*, 19:4967 (1991); Eckert *et al.*, *PCR Methods and Applications*, 1:17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202. The nucleic acid molecules can be amplified using cDNA, mRNA or genomic DNA as a template, cloned into an appropriate vector and characterized by DNA sequence analysis.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics*, 4:560 (1989), Landegren *et al.*, *Science,* 241:1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87:1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be radiolabelled and used as a probe for screening a cDNA library derived from human cells, mRNA in zap express, ZIPLOX or other suitable vector. Corresponding clones can be isolated, DNA can obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art recognized methods to identify the correct reading frame encoding a polypeptide of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of nucleic acid molecules of the present invention can be accomplished using well-known methods that are commercially available. See, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)). Using these or similar methods, the polypeptide and the DNA encoding the polypeptide can be isolated, sequenced and further characterized.

Antisense nucleic acid molecules of the invention can be designed using the nucleotide sequences of SEQ ID NO: 1 and/or the complement of SEQ ID NO: 1, and/or a portion of SEQ ID NO: 1 or the complement of SEQ ID NO: 1, and/or a sequence encoding the amino acid sequence of SEQ ID NO: 2, or encoding a portion of SEQ ID NO: 2, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid molecule (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*,

phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid molecule can be produced biologically using an expression vector into which a nucleic acid molecule has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid molecule will be of an antisense orientation to a target nucleic acid of interest).

In general, the isolated nucleic acid sequences of the invention can be used as molecular weight markers on Southern gels, and as chromosome markers which are labeled to map related gene positions. The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify genetic disorders (*e.g.*, a predisposition for or susceptibility to Parkinson's disease), and as probes, such as to hybridize and discover related DNA sequences or to subtract out known sequences from a sample. The nucleic acid sequences can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using DNA immunization techniques, and as an antigen to raise anti-DNA antibodies or elicit immune responses. Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Additionally, the nucleotide sequences of the invention can be used to identify and express recombinant polypeptides for analysis, characterization or therapeutic use, or as markers for tissues in which the corresponding polypeptide is expressed, either constitutively, during tissue differentiation, or in diseased states. The nucleic acid sequences can additionally be used as reagents in the screening and/or diagnostic assays described herein, and can also be included as components of kits (*e.g.*, reagent kits) for use in the screening and/or diagnostic assays described herein.

Another aspect of the invention pertains to nucleic acid constructs containing a nucleic acid molecule selected from the group consisting of SEQ ID NO: 1 and the complement of SEQ ID NO: 1 (or a portion thereof). Yet another aspect of the

invention pertains to nucleic acid constructs containing a nucleic acid molecule encoding the amino acid sequence of SEQ ID NO: 2. The constructs comprise a vector (*e.g.*, an expression vector) into which a sequence of the invention has been inserted in a sense or antisense orientation. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses) that serve equivalent functions.

Preferred recombinant expression vectors of the invention comprise a nucleic acid molecule of the invention in a form suitable for expression of the nucleic acid molecule in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in

Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed and the level of expression of polypeptide desired. The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides, including fusion polypeptides, encoded by nucleic acid molecules as described herein .

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic or eukaryotic cells, *e.g.*, bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a nucleic acid molecule of the invention can be expressed in bacterial cells (*e.g.*, *E. coli*), insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms

"transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing a foreign nucleic acid molecule (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector as the nucleic acid molecule of the invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid molecule can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a nucleic acid molecule of the invention (*e.g.*, an exogenous RNF 11 gene, or an exogenous nucleic acid encoding an RNF 11 polypeptide) has been introduced. Such host cells can then be

used to create non-human transgenic animals in which exogenous nucleotide sequences have been introduced into the genome or homologous recombinant animals in which endogenous nucleotide sequences have been altered. Such animals are useful for studying the function and/or activity of the nucleotide sequence and polypeptide encoded by the sequence and for identifying and/or evaluating modulators of their activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens and amphibians. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, *Current Opinion in Bio/Technology*, 2:823-829 (1991) and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169. Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.*, *Nature*, 385:810-813 (1997) and PCT Publication Nos. WO 97/07668 and WO 97/07669.

POLYPEPTIDES OF THE INVENTION

The present invention also pertains to isolated polypeptides encoded by RNF 11 ("RNF 11 polypeptides"), and fragments and variants thereof, as well as polypeptides encoded by nucleotide sequences described herein (*e.g.*, other variants). The term "polypeptide" refers to a polymer of amino acids, and not to a specific length; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell (*e.g.*, in a "fusion protein") and still be "isolated" or "purified."

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity. In one embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (*i.e.*, contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

When a polypeptide is recombinantly produced, it can also be substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the polypeptide preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less

than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, a polypeptide of the invention comprises an amino acid sequence encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 and complements and portions thereof.

The polypeptides of the invention also encompass fragments and sequence variants. Variants include a substantially homologous polypeptide encoded by the same genetic locus in an organism, *i.e.*, an allelic variant, as well as other variants. Variants also encompass polypeptides derived from other genetic loci in an organism, but having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO: 1 and complements and portions thereof, or having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence of nucleotide sequences encoding by SEQ ID NO: 1 or variant or portion thereof. Variants also include polypeptides substantially homologous or identical to these polypeptides but derived from another organism, *i.e.*, an ortholog. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by chemical synthesis. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by recombinant methods.

As used herein, two polypeptides (or a region of the polypeptides) are substantially homologous or identical when the amino acid sequences are at least about 45-55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically greater than about 90 or 95% or more homologous or identical. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid molecule hybridizing to SEQ ID NO: 1 or portion thereof, under stringent conditions as more particularly described above, or will be encoded by a nucleic acid molecule hybridizing to a nucleic acid sequence encoding SEQ ID NO: 2, or portion thereof, under stringent conditions as more particularly described thereof.

To determine the percent homology or identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one polypeptide or nucleic acid molecule for optimal alignment with the other polypeptide or nucleic acid molecule). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence, then the molecules are homologous at that position. As used herein, amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity". The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, percent homology equals the number of identical positions/total number of positions times 100).

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by a polypeptide encoded by a nucleic acid molecule of the invention. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science*, 247:1306-1310 (1990).

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these. Further, variant polypeptides can be fully functional or can lack function in one or more activities. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical

regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science,* 244: 1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity *in vitro*, or *in vitro* proliferative activity. Sites that are critical for polypeptide activity can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.,* 224: 899-904 (1992); de Vos *et al. Science,* 255: 306-312 (1992)).

The invention also includes polypeptide fragments of the polypeptides of the invention. Fragments can be derived from a polypeptide encoded by a nucleic acid molecule comprising SEQ ID NO: 1 or a portion thereof and the complements thereof. However, the invention also encompasses fragments of the variants of the polypeptides described herein. As used herein, a fragment comprises at least 6 contiguous amino acids. Useful fragments include those that retain one or more of the biological activities of the polypeptide as well as fragments that can be used as an immunogen to generate polypeptide-specific antibodies.

Biologically active fragments (peptides which are, for example, 6, 9, 12, 15, 16, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) can comprise a domain, segment, or motif that has been identified by analysis of the polypeptide sequence using well-known methods, *e.g.*, signal peptides, extracellular domains, one or more transmembrane segments or loops, ligand binding regions, zinc finger domains, DNA binding domains, acylation sites, glycosylation sites, or phosphorylation sites.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion polypeptides. These comprise a polypeptide of the invention operatively linked to a heterologous protein or polypeptide having an amino acid sequence not substantially homologous to the polypeptide. "Operatively linked" indicates that the polypeptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the polypeptide. In one embodiment the fusion polypeptide does not affect function of the polypeptide *per se*. For example, the fusion polypeptide can be a GST-fusion polypeptide in which the polypeptide sequences are fused to the C-terminus of the GST sequences. Other types of fusion polypeptides include, but are not limited to, enzymatic fusion polypeptides, for example β-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions and Ig fusions. Such fusion polypeptides, particularly poly-His fusions, can facilitate the purification of recombinant polypeptide. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of a polypeptide can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion polypeptide contains a heterologous signal sequence at its N-terminus.

EP-A-0 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.*, *Journal of Molecular Recognition,* 8:52-58 (1995) and Johanson *et al.*, *The Journal of Biological Chemistry*, 270(16): 9459-9471 (1995). Thus, this invention also encompasses soluble fusion polypeptides containing a polypeptide of the

invention and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE).

A chimeric or fusion polypeptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of nucleic acid fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive nucleic acid fragments which can subsequently be annealed and re-amplified to generate a chimeric nucleic acid sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST protein). A nucleic acid molecule encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide.

The isolated polypeptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. In one embodiment, the polypeptide is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the polypeptide expressed in the host cell. The polypeptide can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

In general, polypeptides of the present invention can be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using art-recognized methods. The polypeptides of the present invention can be used to raise antibodies or to elicit an immune response. The polypeptides can also be used as a reagent, *e.g.*, a labeled reagent, in assays to quantitatively determine levels of the polypeptide or a molecule to which it binds (*e.g.*, a receptor or a ligand) in biological fluids. The polypeptides can also be used as markers for cells or tissues in which the corresponding polypeptide is preferentially expressed, either

constitutively, during tissue differentiation, or in a diseased state. The polypeptides can be used to isolate a corresponding binding agent, *e.g.*, receptor or ligand, such as, for example, in an interaction trap assay, and to screen for peptide or small molecule antagonists or agonists of the binding interaction.

ANTIBODIES OF THE INVENTION

Polyclonal and/or monoclonal antibodies that specifically bind one form of the gene product but not to the other form of the gene are also provided. Antibodies are also provided that bind a portion of either variant or the reference gene product that contains the polymorphic site or sites.

In another aspect, the invention provides antibodies to the polypeptides and polypeptide fragments of the invention, *e.g.*, having an amino acid sequence encoded by SEQ ID NO: 2, or a portion thereof, or having an amino acid sequence encoded by a nucleic acid molecule comprising all or a portion of SEQ ID NO: 1 (*e.g.*, SEQ ID NO: 2, or another variant, or portion thereof). The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds an antigen. A molecule that specifically binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind to a polypeptide of the invention. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, *e.g.*, polypeptide of the invention or fragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules directed against the polypeptide can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, *Nature*, 256:495-497 (1975), the human B cell hybridoma technique (Kozbor *et al., Immunol. Today*, 4:72 (1983)), the EBV-hybridoma technique (Cole *et al., Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)) or trioma techniques. The technology for producing hybridomas is well known (see generally Ausubel, *et al.* (Eds.), *Current Protocols in Immunology*, John Wiley & Sons, Inc., New York, NY (2001)). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention (see, *e.g.*, *Current Protocols in Immunology, supra*; Galfre *et al., Nature*, 266:55052 (1977); R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner, *Yale J. Biol. Med.*, 54:387-402 (1981)). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to a polypeptide of the invention can be identified and isolated

by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *Surf*ZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, *Bio/Technology*, 9:1370-1372 (1991); Hay *et al.*, *Hum. Antibod. Hybridomas,* 3:81-85 (1992); Huse *et al.*, *Science,* 246:1275-1281 (1989); Griffiths *et al.*, *EMBO J.,* 12:725-734 (1993).

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

The invention also is intended to cover human antibodies. Their methods for production, isolation purification and use are known to those skilled in the art using standard methodologies.

In general, antibodies of the invention (*e.g.*, a monoclonal antibody) can be used to isolate a polypeptide of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly produced polypeptide expressed in host cells. Moreover, an antibody specific for a polypeptide of the invention can be used to detect the polypeptide (*e.g.*, in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for

example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

DIAGNOSTIC AND SCREENING ASSAYS OF THE INVENTION

The present invention also pertains to diagnostic assays for assessing ring finger protein 11 gene expression, or for assessing activity of RNF 11 polypeptides of the invention. In one embodiment, the assays are used in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with Parkinson's disease, or is at risk for (has a predisposition for or a susceptibility to) developing Parkinson's disease. The invention also provides for prognostic (or predictive) assays for determining whether an individual is susceptible to developing Parkinson's disease. For example, mutations in the gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of symptoms associated with Parkinson's disease. Another aspect of the invention pertains to assays for monitoring the influence of agents (*e.g.*, drugs, compounds or other agents) on the gene expression or activity of polypeptides of the invention, as well as to assays for identifying agents which bind to RNF 11 polypeptides. These and other assays and agents are described in further detail in the following sections.

DIAGNOSTIC ASSAYS

Nucleic acids, probes, primers, polypeptides and antibodies to the RNF 11 protein can be used in methods of diagnosis of a susceptibility to Parkinson's disease, as well as in kits useful for diagnosis of a susceptibility to Parkinson's disease.

In one embodiment of the invention, diagnosis of a susceptibility to Parkinson's disease is made by detecting a polymorphism in *PARK8*. The polymorphism can be a mutation in *PARK8*, such as the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift mutation; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of the gene; duplication of all or a part of the gene; transposition of all or a part of the gene; or rearrangement of all or a part of the gene. More than one such mutation may be present in a single gene. Such sequence changes cause a mutation in the polypeptide encoded by *PARK8*. For example, if the mutation is a frame shift mutation, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with a susceptibility to Parkinson's disease can be a synonymous mutation in one or more nucleotides (*i.e.*, a mutation that does not result in a change in the polypeptide encoded by *PARK8*). Such a polymorphism may alter binding sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the gene. *PARK8* that has any of the mutations described above is referred to herein as a "mutant or variant gene."

In a first method of diagnosing a susceptibility to Parkinson's disease, hybridization methods, such as Southern analysis, Northern analysis, or *in situ* hybridizations, can be used (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements through 2001). For

example, a biological sample from a test subject (a "test sample") of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having, being susceptible to or predisposed for, or carrying a defect for, Parkinson's disease (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source which contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a polymorphism in *PARK8* is present, and/or to determine which variant(s) encoded by *PARK8* is present. The presence of the polymorphism or variant(s) can be indicated by hybridization of the gene in the genomic DNA, RNA, or cDNA to a nucleic acid probe. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe can contain at least one polymorphism in *PARK8* or contains a nucleic acid encoding a particular variant of *PARK8*. The probe can be any of the nucleic acid molecules described above (*e.g.*, the gene, a fragment, a vector comprising the gene, a probe or primer, etc.).

To diagnose a susceptibility to Parkinson's disease, a hybridization sample is formed by contacting the test sample containing *PARK8*, with at least one nucleic acid probe. A preferred probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can be all or a portion of SEQ ID NO: 1, or the complement of SEQ ID NO: 1, or a portion thereof; or can be a nucleic acid encoding all or a portion of SEQ ID NO: 2. Other suitable probes for use in the diagnostic assays of the invention are described above (see. *e.g.*, probes and primers discussed under the heading, "Nucleic Acids of the Invention"). In a preferred embodiment, the probe

can be all or a portion of the promoter region of the PARK8 gene or complement comprising at least one of the novel SNPs in the promoter region or the SNP proximate to the ATG start codon.

The hybridization sample is maintained under conditions which are sufficient to allow specific hybridization of the nucleic acid probe to *PARK8*. "Specific hybridization", as used herein, indicates exact hybridization (*e.g.*, with no mismatches). Specific hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and *PARK8* in the test sample, then *PARK8* has the polymorphism, or is the variant, that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a polymorphism in *PARK8*, or of the presence of a particular variant encoded by *PARK8*, and is therefore diagnostic for a susceptibility to Parkinson's disease.

In Northern analysis (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, *supra*), the hybridization methods described above are used to identify the presence of a polymorphism or of a particular variant, associated with a susceptibility to Parkinson's disease. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a polymorphism in *PARK8*, or of the presence of a particular variant encoded by *PARK8*, and is therefore diagnostic for a susceptibility to Parkinson's disease.

For representative examples of use of nucleic acid probes, see, for example, U.S. Patents No. 5,288,611 and 4,851,330.

Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA

mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P.E. *et al., Bioconjugate Chemistry*, 1994, 5, American Chemical Society, p. 1 (1994). The PNA probe can be designed to specifically hybridize to a gene having a polymorphism associated with a susceptibility to Parkinson's disease. Hybridization of the PNA probe to *PARK8* is diagnostic for a susceptibility to Parkinson's disease.

In another method of the invention, mutation analysis by restriction digestion can be used to detect a mutant gene, or genes containing a polymorphism(s), if the mutation or polymorphism in the gene results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify *PARK8* (and, if necessary, the flanking sequences) in the test sample of genomic DNA from the test individual. RFLP analysis is conducted as described (see Current Protocols in Molecular Biology, *supra*). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the mutation or polymorphism in *PARK8*, and therefore indicates the presence or absence of this susceptibility to Parkinson's disease.

Sequence analysis can also be used to detect specific polymorphisms in *PARK8*. A test sample of DNA or RNA is obtained from the test individual. PCR or other appropriate methods can be used to amplify the gene, and/or its flanking sequences, if desired. The sequence of *PARK8*, or a fragment of the gene, or cDNA, or fragment of the cDNA, or mRNA, or fragment of the mRNA, is determined, using standard methods. The sequence of the gene, gene fragment, cDNA, cDNA fragment, mRNA, or mRNA fragment is compared with the known nucleic acid sequence of the gene, cDNA (*e.g.*, SEQ ID NO: 1, or a nucleic acid sequence encoding the protein of SEQ ID NO: 2, or a fragment thereof) or mRNA, as appropriate. The presence of a polymorphism in *PARK8* indicates that the individual has a susceptibility to Parkinson's disease.

Allele-specific oligonucleotides can also be used to detect the presence of a polymorphism in *PARK8*, through the use of dot-blot hybridization of amplified

oligonucleotides with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki, R. *et al.*, *Nature (London)* 324:163-166 (1986)). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, preferably approximately 15-30 base pairs, that specifically hybridizes to *PARK8*, and that contains a polymorphism associated with a susceptibility to Parkinson's disease. An allele-specific oligonucleotide probe that is specific for particular polymorphisms in *PARK8* can be prepared, using standard methods (see Current Protocols in Molecular Biology, *supra*). To identify polymorphisms in the gene that are associated with a susceptibility to Parkinson's disease, a test sample of DNA is obtained from the individual. PCR can be used to amplify all or a fragment of *PARK8*, and its flanking sequences. The DNA containing the amplified *PARK8* (or fragment of the gene) is dot-blotted, using standard methods (see Current Protocols in Molecular Biology, *supra*), and the blot is contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified *PARK8* is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a polymorphism in *PARK8*, and is therefore indicative of a susceptibility to Parkinson's disease.

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual, can be used to identify polymorphisms in *PARK8*. For example, in one embodiment, an oligonucleotide array can be used. Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "Genechips™," have been generally described in the art, for example, U.S. Pat. No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092. These arrays can generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods. See Fodor *et al.*, Science, 251:767-777 (1991), Pirrung *et al.*, U.S. Pat. No. 5,143,854 (see also PCT Application No. WO 90/15070) and Fodor *et al.*, PCT Publication No. WO

92/10092 and U.S. Pat. No. 5,424,186, the entire teachings of each of which are incorporated by reference herein. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, *e.g.*, U.S. Pat. Nos. 5,384,261, the entire teachings of which are incorporated by reference herein.

Once an oligonucleotide array is prepared, a nucleic acid of interest is hybridized with the array and scanned for polymorphisms. Hybridization and scanning are generally carried out by methods described herein and also in, *e.g.*, Published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Pat. No. 5,424,186, the entire teachings of which are incorporated by reference herein. In brief, a target nucleic acid sequence which includes one or more previously identified polymorphic markers is amplified by well known amplification techniques, *e.g.*, PCR. Typically, this involves the use of primer sequences that are complementary to the two strands of the target sequence both upstream and downstream from the polymorphism. Asymmetric PCR techniques may also be used. Amplified target, generally incorporating a label, is then hybridized with the array under appropriate conditions. Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data obtained from the scan is typically in the form of fluorescence intensities as a function of location on the array.

Although primarily described in terms of a single detection block, *e.g.*, for detection of a single polymorphism, arrays can include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms. In alternate arrangements, it will generally be understood that detection blocks may be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions may be used during the hybridization of the target to the array. For example, it may often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of hybridization conditions for each situation.

Additional description of use of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patents 5,858,659 and 5,837,832,

the entire teachings of which are incorporated by reference herein.

Other methods of nucleic acid analysis can be used to detect polymorphisms in *PARK8* or variants encoded by *PARK8*. Representative methods include direct manual sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA*, 81:1991-1995, (1988); Sanger, F. *et al. Proc. Natl. Acad. Sci.,* 74:5463-5467 (1977); Beavis *et al.*, U.S. Pat. No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield, V.C. *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 232-236 (1991)), mobility shift analysis (Orita, M. *et al., Proc. Natl. Acad. Sci. USA*, 86: 2766-2770 (1989)), restriction enzyme analysis (Flavell *et al. Cell* 15: 25 (1978); Geever, *et al., Proc. Natl. Acad. Sci. USA* 78:5081); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 4397-4401 (1985)); RNase protection assays (Myers, R.M. *et al. Science* 230: 1242 (1985)); use of polypeptides which recognize nucleotide mismatches, such as *E. coli* mutS protein; allele-specific PCR, for example.

In another embodiment of the invention, diagnosis of a susceptibility to Parkinson's disease can also be made by examining expression and/or composition of an *PARK8* polypeptide, by a variety of methods, including enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. A test sample from an individual is assessed for the presence of an alteration in the expression and/or an alteration in composition of the polypeptide encoded by *PARK8*, or for the presence of a particular variant encoded by *PARK8*. An alteration in expression of a polypeptide encoded by *PARK8* can be, for example, an alteration in the quantitative polypeptide expression (*i.e.*, the amount of polypeptide produced); an alteration in the composition of a polypeptide encoded by *PARK8* is an alteration in the qualitative polypeptide expression (*e.g.*, expression of a mutant *PARK8* polypeptide or of a different variant). In a preferred embodiment, diagnosis of a susceptibility to Parkinson's disease is made by detecting a particular variant encoded by *PARK8*, or a particular pattern of variants.

Both quantitative and qualitative alterations can also be present. An "alteration" in the polypeptide expression or composition, as used herein, refers to an alteration in expression or composition in a test sample, as compared with the expression or composition of polypeptide by *PARK8* in a control sample. A control sample is a sample that corresponds to the test sample (*e.g.*, is from the same type of cells), and is from an individual who is not affected by Parkinson's disease. An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, is indicative of a susceptibility to Parkinson's disease. Similarly, the presence of one or more different variants in the test sample, or the presence of significantly different amounts of different variants in the test sample, as compared with the control sample, is indicative of a susceptibility to Parkinson's disease. Various means of examining expression or composition of the polypeptide encoded by *PARK8* can be used, including spectroscopy, colorimetry, electrophoresis, isoelectric focusing, and immunoassays (*e.g.*, David *et al.*, U.S. Pat. No. 4,376,110) such as immunoblotting (see also *Current Protocols in Molecular Biology*, *supra*., particularly Chapter 10). For example, in one embodiment, an antibody capable of binding to the polypeptide (*e.g.*, as described above), preferably an antibody with a detectable label, can be used. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

Western blotting analysis, using an antibody as described above that specifically binds to a polypeptide encoded by a mutant *PARK8*, or an antibody that specifically binds to a polypeptide encoded by a non-mutant gene, or an antibody that specifically binds to a particular variant encoded by *PARK8*, can be used to

identify the presence in a test sample of a particular variant or of a polypeptide encoded by a polymorphic or mutant *PARK8*, or the absence in a test sample of a particular variant or of a polypeptide encoded by a non-polymorphic or non-mutant gene. The presence of a polypeptide encoded by a polymorphic or mutant gene, or the absence of a polypeptide encoded by a non-polymorphic or non-mutant gene, is diagnostic for a susceptibility to Parkinson's disease, as is the presence (or absence) of particular variants encoded by the *PARK8* gene.

In one embodiment of this method, the level or amount of polypeptide encoded by *PARK8* in a test sample is compared with the level or amount of the polypeptide encoded by *PARK8* in a control sample. A level or amount of the polypeptide in the test sample that is higher or lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the polypeptide encoded by *PARK8*, and is diagnostic for a susceptibility to Parkinson's disease. Alternatively, the composition of the polypeptide encoded by *PARK8* in a test sample is compared with the composition of the polypeptide encoded by *PARK8* in a control sample. A difference in the composition of the polypeptide in the test sample, as compared with the composition of the polypeptide in the control sample (*e.g.*, the presence of different variants), is diagnostic for a susceptibility to Parkinson's disease. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample. A difference in the amount or level of the polypeptide in the test sample, compared to the control sample; a difference in composition in the test sample, compared to the control sample; or both a difference in the amount or level, and a difference in the composition, is indicative of a susceptibility to Parkinson's disease.

Kits (*e.g.*, reagent kits) useful in the methods of diagnosis comprise components useful in any of the methods described herein, including for example, hybridization probes or primers as described herein (*e.g.*, labeled probes or primers), reagents for detection of labeled molecules, restriction enzymes (*e.g.*, for RFLP analysis), allele-specific oligonucleotides, antibodies which bind to mutant or to non-mutant (native) *PARK8* polypeptide, means for amplification of nucleic acids

comprising *PARK8*, or means for analyzing the nucleic acid sequence of *PARK8* or for analyzing the amino acid sequence of an *PARK8* polypeptide, etc.

SCREENING ASSAYS AND AGENTS IDENTIFIED THEREBY

The invention provides methods (also referred to herein as "screening assays") for identifying the presence of a nucleotide that hybridizes to a nucleic acid of the invention, as well as for identifying the presence of a polypeptide encoded by a nucleic acid of the invention. In one embodiment, the presence (or absence) of a nucleic acid molecule of interest (*e.g.*, a nucleic acid that has significant homology with a nucleic acid of *PARK8*) in a sample can be assessed by contacting the sample with a nucleic acid comprising a nucleic acid of the invention (*e.g.*, a nucleic acid having the sequence of SEQ ID NO: 1 which may optionally comprise at least one polymorphism, or the complement thereof, or a nucleic acid encoding an amino acid having the sequence of SEQ ID NO: 2 or a fragment or variant of such nucleic acids), under stringent conditions as described above, and then assessing the sample for the presence (or absence) of hybridization. In a preferred embodiment, high stringency conditions are conditions appropriate for selective hybridization. In another embodiment, a sample containing the nucleic acid molecule of interest is contacted with a nucleic acid containing a contiguous nucleotide sequence (*e.g.*, a primer or a probe as described above) that is at least partially complementary to a part of the nucleic acid molecule of interest (*e.g.*, a *PARK8* nucleic acid), and the contacted sample is assessed for the presence or absence of hybridization. In a preferred embodiment, the nucleic acid containing a contiguous nucleotide sequence is completely complementary to a part of the nucleic acid molecule of *PARK8*.

In any of these embodiments, all or a portion of the nucleic acid of interest can be subjected to amplification prior to performing the hybridization.

In another embodiment, the presence (or absence) of a polypeptide of *PARK8*, such as a polypeptide of the invention or a fragment or variant thereof, in a sample can be assessed by contacting the sample with an antibody that specifically hybridizes to the polypeptide of *PARK8* (*e.g.*, an antibody such as those described

above), and then assessing the sample for the presence (or absence) of binding of the antibody to the polypeptide of *PARK8*.

In another embodiment, the invention provides methods for identifying agents (*e.g.*, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes which alter (*e.g.*, increase or decrease) the activity of the polypeptides described herein, or which otherwise interact with the polypeptides herein. For example, such agents can be agents which bind to polypeptides described herein (*e.g.*, *PARK8* binding agents); which have a stimulatory or inhibitory effect on, for example, activity of polypeptides of the invention; or which change (*e.g.*, enhance or inhibit) the ability of the polypeptides of the invention to interact with *PARK8* binding agents (*e.g.*, receptors or other binding agents); or which alter posttranslational processing of the *PARK8* polypeptide (*e.g.*, agents that alter proteolytic processing to direct the polypeptide from where it is normally synthesized to another location in the cell, such as the cell surface; agents that alter proteolytic processing such that more polypeptide is released from the cell, etc.

In one embodiment, the invention provides assays for screening candidate or test agents that bind to or modulate the activity of polypeptides described herein (or biologically active portion(s) thereof), as well as agents identifiable by the assays. Test agents can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S., *Anticancer Drug Des.*, 12:145 (1997)).

In one embodiment, to identify agents which alter the activity of a *PARK8* polypeptide, a cell, cell lysate, or solution containing or expressing a *PARK8* polypeptide (*e.g.*, SEQ ID NO: 2 or another variant encoded by *PARK8*), or a fragment or derivative thereof (as described above), can be contacted with an agent

to be tested; alternatively, the polypeptide can be contacted directly with the agent to be tested. The level (amount) of *PARK8* activity is assessed (*e.g.*, the level (amount) of *PARK8* activity is measured, either directly or indirectly), and is compared with the level of activity in a control (*i.e.*, the level of activity of the *PARK8* polypeptide or active fragment or derivative thereof in the absence of the agent to be tested). If the level of the activity in the presence of the agent differs, by an amount that is statistically significant, from the level of the activity in the absence of the agent, then the agent is an agent that alters the activity of *PARK8* polypeptide. An increase in the level of *PARK8* activity relative to a control, indicates that the agent is an agent that enhances (is an agonist of) *PARK8* activity. Similarly, a decrease in the level of *PARK8* activity relative to a control, indicates that the agent is an agent that inhibits (is an antagonist of) *PARK8* activity. In another embodiment, the level of activity of a *PARK8* polypeptide or derivative or fragment thereof in the presence of the agent to be tested, is compared with a control level that has previously been established. A level of the activity in the presence of the agent that differs from the control level by an amount that is statistically significant indicates that the agent alters *PARK8* activity.

The present invention also relates to an assay for identifying agents which alter the expression of the *PARK8* gene (*e.g.*, antisense nucleic acids, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) which alter (*e.g.*, increase or decrease) expression (*e.g.*, transcription or translation) of the gene or which otherwise interact with the nucleic acids described herein, as well as agents identifiable by the assays. For example, a solution containing a nucleic acid encoding *PARK8* polypeptide (*e.g.*, *PARK8* gene) can be contacted with an agent to be tested. The solution can comprise, for example, cells containing the nucleic acid or cell lysate containing the nucleic acid; alternatively, the solution can be another solution which comprises elements necessary for transcription/translation of the nucleic acid. Cells not suspended in solution can also be employed, if desired. The level and/or pattern of *PARK8* expression (*e.g.*, the level and/or pattern of mRNA or of protein expressed, such as the level and/or pattern of different variants) is

assessed, and is compared with the level and/or pattern of expression in a control (*i.e.*, the level and/or pattern of the *PARK8* expression in the absence of the agent to be tested). If the level and/or pattern in the presence of the agent differs, by an amount or in a manner that is statistically significant, from the level and/or pattern in the absence of the agent, then the agent is an agent that alters the expression of *PARK8*. Enhancement of *PARK8* expression indicates that the agent is an agonist of *PARK8* activity. Similarly, inhibition of *PARK8* expression indicates that the agent is an antagonist of *PARK8* activity. In another embodiment, the level and/or pattern of *PARK8* polypeptide(s) (*e.g.*, different variants) in the presence of the agent to be tested, is compared with a control level and/or pattern that has previously been established. A level and/or pattern in the presence of the agent that differs from the control level and/or pattern by an amount or in a manner that is statistically significant indicates that the agent alters *PARK8* expression.

In another embodiment of the invention, agents which alter the expression of the *PARK8* gene or which otherwise interact with the nucleic acids described herein, can be identified using a cell, cell lysate, or solution containing a nucleic acid encoding the promoter region of the *PARK8* gene operably linked to a reporter gene. After contact with an agent to be tested, the level of expression of the reporter gene (*e.g.*, the level of mRNA or of protein expressed) is assessed, and is compared with the level of expression in a control (*i.e.*, the level of the expression of the reporter gene in the absence of the agent to be tested). If the level in the presence of the agent differs, by an amount or in a manner that is statistically significant, from the level in the absence of the agent, then the agent is an agent that alters the expression of *PARK8*, as indicated by its ability to alter expression of a gene that is operably linked to the *PARK8* gene promoter. Enhancement of the expression of the reporter indicates that the agent is an agonist of *PARK8* activity. Similarly, inhibition of the expression of the reporter indicates that the agent is an antagonist of *PARK8* activity. In another embodiment, the level of expression of the reporter in the presence of the agent to be tested, is compared with a control level that has previously been established. A level in the presence of the agent that differs from the control level

by an amount or in a manner that is statistically significant indicates that the agent alters *PARK8* expression.

Agents which alter the amounts of different variants encoded by *PARK8* (*e.g.*, an agent which enhances activity of a first variant, and which inhibits activity of a second variant), as well as agents which are agonists of activity of a first variant and antagonists of activity of a second variant, can easily be identified using these methods described above.

In other embodiments of the invention, assays can be used to assess the impact of a test agent on the activity of a polypeptide in relation to a *PARK8* binding agent. For example, a cell that expresses a compound that interacts with *PARK8* (herein referred to as a "*PARK8* binding agent", which can be a polypeptide or other molecule that interacts with *PARK8*, such as a receptor) is contacted with *PARK8* in the presence of a test agent, and the ability of the test agent to alter the interaction between *PARK8* and the *PARK8* binding agent is determined. Alternatively, a cell lysate or a solution containing the *PARK8* binding agent, can be used. An agent which binds to *PARK8* or the *PARK8* binding agent can alter the interaction by interfering with, or enhancing the ability of *PARK8* to bind to, associate with, or otherwise interact with the *PARK8* binding agent. Determining the ability of the test agent to bind to *PARK8* or an *PARK8* binding agent can be accomplished, for example, by coupling the test agent with a radioisotope or enzymatic label such that binding of the test agent to the polypeptide can be determined by detecting the labeled with ^{125}I, ^{35}S, ^{14}C or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test agents can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. It is also within the scope of this invention to determine the ability of a test agent to interact with the polypeptide without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test agent with *PARK8* or a *PARK8* binding agent without the labeling of either the test agent, *PARK8*, or the *PARK8* binding agent. McConnell, H.M. *et al.*, *Science*, 257:

1906-1912 (1992). As used herein, a "microphysiometer" (*e.g.*, Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and polypeptide. See the Examples Section for a discussion of know *PARK8* binding partners. Thus, these receptors can be used to screen for compounds that are *PARK8* receptor agonists for use in treating Parkinson's disease or *PARK8* receptor antagonists for studying Parkinson's disease. The linkage data provided herein, for the first time, provides such correction to Parkinson's disease. Drugs could be designed to regulate *PARK8* receptor activation which in turn can be used to regulate signaling pathways and transcription events of genes downstream.

In another embodiment of the invention, assays can be used to identify polypeptides that interact with one or more *PARK8* polypeptides, as described herein. For example, a yeast two-hybrid system such as that described by Fields and Song (Fields, S. and O. Song, *Nature* 340: 245-246 (1989)) can be used to identify polypeptides that interact with one or more *PARK8* polypeptides. In such a yeast two-hybrid system, vectors are constructed based on the flexibility of a transcription factor which has two functional domains (a DNA binding domain and a transcription activation domain). If the two domains are separated but fused to two different proteins that interact with one another, transcriptional activation can be achieved, and transcription of specific markers (*e.g.*, nutritional markers such as His and Ade, or color markers such as lacZ) can be used to identify the presence of interaction and transcriptional activation. For example, in the methods of the invention, a first vector is used which includes a nucleic acid encoding a DNA binding domain and also an *PARK8* polypeptide, variant, or fragment or derivative thereof, and a second vector is used which includes a nucleic acid encoding a transcription activation domain and also a nucleic acid encoding a polypeptide which potentially may interact with the *PARK8* polypeptide, variant, or fragment or derivative thereof (*e.g.*, a *PARK8* polypeptide binding agent or receptor). Incubation of yeast containing the first vector and the second vector under appropriate conditions (*e.g.*, mating conditions such as used in the Matchmaker™ system from Clontech) allows

identification of colonies which express the markers of *PARK8*. These colonies can be examined to identify the polypeptide(s) which interact with the *PARK8* polypeptide or fragment or derivative thereof. Such polypeptides may be useful as agents which alter the activity of expression of an *PARK8* polypeptide, as described above.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either *PARK8*, the *PARK8* binding agent, or other components of the assay on a solid support, in order to facilitate separation of complexed from uncomplexed forms of one or both of the polypeptides, as well as to accommodate automation of the assay. Binding of a test agent to the polypeptide, or interaction of the polypeptide with a binding agent in the presence and absence of a test agent, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein (*e.g.*, a glutathione-S-transferase fusion protein) can be provided which adds a domain that allows *PARK8* or a *PARK8* binding agent to be bound to a matrix or other solid support.

In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell, cell lysate, or solution containing a nucleic acid encoding *PARK8* is contacted with a test agent and the expression of appropriate mRNA or polypeptide (*e.g.*, variant(s)) in the cell, cell lysate, or solution, is determined. The level of expression of appropriate mRNA or polypeptide(s) in the presence of the test agent is compared to the level of expression of mRNA or polypeptide(s) in the absence of the test agent. The test agent can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or polypeptide is greater (statistically significantly greater) in the presence of the test agent than in its absence, the test agent is identified as a stimulator or enhancer of the mRNA or polypeptide expression. Alternatively, when expression of the mRNA or polypeptide is less (statistically significantly less) in the presence of the test agent than in its absence, the test agent is identified as an inhibitor of the mRNA or polypeptide expression. The level of

mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting mRNA or polypeptide.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a test agent that is a modulating agent, an antisense nucleic acid molecule, a specific antibody, or a polypeptide-binding agent) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein. In addition, an agent identified as described herein can be used to alter activity of a polypeptide encoded by *PARK8*, or to alter expression of *PARK8*, by contacting the polypeptide or the gene (or contacting a cell comprising the polypeptide or the gene) with the agent identified as described herein.

PHARMACEUTICAL COMPOSITIONS

The present invention also pertains to pharmaceutical compositions comprising nucleic acids described herein, particularly nucleotides encoding the polypeptides described herein; comprising polypeptides described herein (*e.g.*, SEQ ID NO: 2, and/or other variants encoded by *PARK8*); and/or comprising an agent that alters (*e.g.*, enhances or inhibits) *PARK8* expression or *PARK8* polypeptide activity as described herein. For instance, a polypeptide, protein (*e.g.*, an *PARK8* receptor), fragment, fusion protein or prodrug thereof, or a nucleotide or nucleic acid construct (vector) comprising a nucleotide of the present invention, an agent that alters *PARK8* polypeptide activity, an agent that alters Parkinson's disease gene expression, or an *PARK8* binding agent or binding partner, can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical

composition. The carrier and composition can be sterile. The formulation should suit the mode of administration.

Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (*e.g.*, NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrolidone, etc., as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents, *e.g.*, lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances and the like which do not deleteriously react with the active agents.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrollidone, sodium saccharine, cellulose, magnesium carbonate, etc.

Methods of introduction of these compositions include, but are not limited to, intradermal, intramuscular, intraperitoneal, intraocular, intravenous, subcutaneous, topical, oral and intranasal. Other suitable methods of introduction can also include gene therapy (as described below), rechargeable or biodegradable devices, particle acceleration devises ("gene guns") and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents.

The composition can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human beings. For example, compositions for intravenous administration typically are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may

also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

For topical application, nonsprayable forms, viscous to semi-solid or solid forms comprising a carrier compatible with topical application and having a dynamic viscosity preferably greater than water, can be employed. Suitable formulations include but are not limited to solutions, suspensions, emulsions, creams, ointments, powders, enemas, lotions, sols, liniments, salves, aerosols, etc., which are, if desired, sterilized or mixed with auxiliary agents, *e.g.*, preservatives, stabilizers, wetting agents, buffers or salts for influencing osmotic pressure, etc. The agent may be incorporated into a cosmetic formulation. For topical application, also suitable are sprayable aerosol preparations wherein the active ingredient, preferably in combination with a solid or liquid inert carrier material, is packaged in a squeeze bottle or in admixture with a pressurized volatile, normally gaseous propellant, *e.g.*, pressurized air.

Agents described herein can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The agents are administered in a therapeutically effective amount. The amount of agents which will be therapeutically effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in*

vitro or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the symptoms of Parkinson's disease, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use of sale for human administration. The pack or kit can be labeled with information regarding mode of administration, sequence of drug administration (*e.g.*, separately, sequentially or concurrently), or the like. The pack or kit may also include means for reminding the patient to take the therapy. The pack or kit can be a single unit dosage of the combination therapy or it can be a plurality of unit dosages. In particular, the agents can be separated, mixed together in any combination, present in a single vial or tablet. Agents assembled in a blister pack or other dispensing means is preferred. For the purpose of this invention, unit dosage is intended to mean a dosage that is dependent on the individual pharmacodynamics of each agent and administered in FDA approved dosages in standard time courses.

METHODS OF THERAPY

The present invention also pertains to methods of treatment (prophylactic and/or therapeutic) for Parkinson's disease, using a *PARK8* therapeutic agent. A "*PARK8* therapeutic agent" is an agent that alters (*e.g.*, enhances or inhibits) *PARK8* polypeptide activity and/or *PARK8* gene expression, as described herein (*e.g.*, a *PARK8* agonist or antagonist). *PARK8* therapeutic agents can alter *PARK8* polypeptide activity or gene expression by a variety of means, such as, for example, by providing additional *PARK8* polypeptide or by upregulating the transcription or

translation of the *PARK8* gene; by altering posttranslational processing of the *PARK8* polypeptide; by altering transcription of *PARK8* variants; or by interfering with *PARK8* polypeptide activity (*e.g.*, by binding to a *PARK8* polypeptide), or by downregulating the transcription or translation of the *PARK8* gene. Representative *PARK8* therapeutic agents include the following: nucleic acids or fragments or derivatives thereof described herein, particularly nucleotides encoding the polypeptides described herein and vectors comprising such nucleic acids (*e.g.*, a gene, cDNA, and/or mRNA, such as a nucleic acid encoding a *PARK8* polypeptide or active fragment or derivative thereof, or an oligonucleotide); polypeptides described herein and/or other variants encoded by *PARK8*, or fragments or derivatives thereof; other polypeptides (*e.g.*, *PARK8* receptors); *PARK8* binding agents; peptidomimetics; fusion proteins or prodrugs thereof; antibodies (*e.g.*, an antibody to a mutant *PARK8* polypeptide, or an antibody to a non-mutant *PARK8* polypeptide, or an antibody to a particular variant encoded by *PARK8*, as described above); ribozymes; other small molecules; and other agents that alter (*e.g.*, enhance or inhibit) *PARK8* gene expression or polypeptide activity, or that regulate transcription of *PARK8* variants (*e.g.*, agents that affect which variants are expressed, or that affect the amount of each variant that is expressed). More than one *PARK8* therapeutic agent can be used concurrently, if desired.

The *PARK8* therapeutic agent that is a nucleic acid is used in the treatment of Parkinson's disease. The term "treatment", as used herein, refers not only to ameliorating symptoms associated with the disease, but also preventing or delaying the onset of the disease, and also lessening the severity or frequency of symptoms of the disease. The therapy is designed to alter (*e.g.*, inhibit or enhance), replace or supplement activity of a *PARK8* polypeptide in an individual. For example, a *PARK8* therapeutic agent can be administered in order to upregulate or increase the expression or availability of the *PARK8* gene or of specific variants of *PARK8*, or, conversely, to downregulate or decrease the expression or availability of the *PARK8* gene or specific variants of *PARK8*. Upregulation or increasing expression or availability of a native *PARK8* gene or of a particular variant could interfere with or compensate for the expression or activity of a defective gene or another variant;

downregulation or decreasing expression or availability of a native *PARK8* gene or of a particular variant could minimize the expression or activity of a defective gene or the particular variant and thereby minimize the impact of the defective gene or the particular variant.

The *PARK8* therapeutic agent(s) are administered in a therapeutically effective amount (*i.e.*, an amount that is sufficient to treat the disease, such as by ameliorating symptoms associated with the disease, preventing or delaying the onset of the disease, and/or also lessening the severity or frequency of symptoms of the disease). The amount which will be therapeutically effective in the treatment of a particular individual's disorder or condition will depend on the symptoms and severity of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

In one embodiment, a nucleic acid of the invention can be used, either alone or in a pharmaceutical composition as described above. For example, *PARK8* or a cDNA encoding the *PARK8* polypeptide, either by itself or included within a vector, can be introduced into cells (either *in vitro* or *in vivo*) such that the cells produce native *PARK8* polypeptide. If necessary, cells that have been transformed with the gene or cDNA or a vector comprising the gene or cDNA can be introduced (or re-introduced) into an individual affected with the disease. Thus, cells which, in nature, lack native *PARK8* expression and activity, or have mutant *PARK8* expression and activity, or have expression of a disease-associated *PARK8* variant, can be engineered to express *PARK8* polypeptide or an active fragment of the *PARK8* polypeptide (or a different variant of *PARK8* polypeptide). In a preferred embodiment, nucleic acid encoding the *PARK8* polypeptide, or an active fragment or derivative thereof, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells in an animal. Other

gene transfer systems, including viral and nonviral transfer systems, can be used. Alternatively, nonviral gene transfer methods, such as calcium phosphate coprecipitation, mechanical techniques (*e.g.*, microinjection); membrane fusion-mediated transfer via liposomes; or direct DNA uptake, can also be used.

Alternatively, in another embodiment of the invention, a nucleic acid of the invention; a nucleic acid complementary to a nucleic acid of the invention; or a portion of such a nucleic acid (*e.g.*, an oligonucleotide as described below), can be used in "antisense" therapy, in which a nucleic acid (*e.g.*, an oligonucleotide) which specifically hybridizes to the mRNA and/or genomic DNA of *PARK8* is administered or generated *in situ*. The antisense nucleic acid that specifically hybridizes to the mRNA and/or DNA inhibits expression of the *PARK8* polypeptide, *e.g.*, by inhibiting translation and/or transcription. Binding of the antisense nucleic acid can be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interaction in the major groove of the double helix.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid as described above. When the plasmid is transcribed in the cell, it produces RNA which is complementary to a portion of the mRNA and/or DNA which encodes *PARK8* polypeptide. Alternatively, the antisense construct can be an oligonucleotide probe which is generated *ex vivo* and introduced into cells; it then inhibits expression by hybridizing with the mRNA and/or genomic DNA of *PARK8*. In one embodiment, the oligonucleotide probes are modified oligonucleotides which are resistant to endogenous nucleases, *e.g.* exonucleases and/or endonucleases, thereby rendering them stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Pat. Nos. 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy are also described, for example, by Van der Krol *et al.*, *Biotechniques*, 6:958-976 (1988); and Stein *et al.*, *Cancer Res.*, 48:2659-2668 (1988). With respect to antisense DNA,

oligodeoxyribonucleotides derived from the translation initiation site, *e.g.* between the -10 and +10 regions of *PARK8* sequence, are preferred.

To perform antisense therapy, oligonucleotides (mRNA, cDNA or DNA) are designed that are complementary to mRNA encoding *PARK8*. The antisense oligonucleotides bind to *PARK8* mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, indicates that a sequence has sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid, as described in detail above. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures.

The oligonucleotides used in antisense therapy can be DNA, RNA, or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotides can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotides can include other appended groups such as peptides (*e.g.* for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al. Proc. Natl. Acad. Sci. USA*, 86:6553-6556 (1989); Lemaitre *et al.*, *Proc. Natl. Acad Sci. USA*, 84:648-652 (1987); PCT International Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT International Publication No. W089/10134), or hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al. BioTechniques*, 6:958-976 (1988)) or intercalating agents. See, *e.g.*, Zon,, *Pharm. Res.*, 5:539-549 (1988). To this end, the oligonucleotide may be conjugated to another molecule (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent).

The antisense molecules are delivered to cells which express *PARK8 in vivo*. A number of methods can be used for delivering antisense DNA or RNA to cells; *e.g.*, antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (*e.g.*, antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systematically. Alternatively, in a preferred embodiment, a recombinant DNA construct is utilized in which the antisense oligonucleotide is placed under the control of a strong promoter (*e.g.*, pol III or pol II). The use of such a construct to transfect target cells in the patient results in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous *PARK8* transcripts and thereby prevent translation of the *PARK8* mRNA. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art and described above. For example, a plasmid, cosmid, YAC or viral vector can be used to prepare the recombinant DNA construct which can be introduced directly into the tissue site. Alternatively, viral vectors can be used which selectively infect the desired tissue, in which case administration may be accomplished by another route (*e.g.*, systematically).

Endogenous *PARK8* expression can also be reduced by inactivating or "knocking out" *PARK8* or its promoter using targeted homologous recombination (*e.g.*, see Smithies *et al.*, *Nature*, 317: 230-234 (1985); Thomas and Capecchi, *Cell*, 51:503-512 (1987); Thompson *et al.*, *Cell*, 5:313-321 (1989)). For example, a mutant, non-functional *PARK8* (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous *PARK8* (either the coding regions or regulatory regions of *PARK8*) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express *PARK8 in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of *PARK8*. The recombinant DNA constructs can be directly

administered or targeted to the required site *in vivo* using appropriate vectors, as described above. Alternatively, expression of non-mutant *PARK8* can be increased using a similar method: targeted homologous recombination can be used to insert a DNA construct comprising a non-mutant, functional *PARK8*, or a portion thereof, in place of a mutant *PARK8* in the cell, as described above. In another embodiment, targeted homologous recombination can be used to insert a DNA construct comprising a nucleic acid that encodes a *PARK8* polypeptide variant that differs from that present in the cell.

Alternatively, endogenous *PARK8* expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of *PARK8* (*i.e.*, the *PARK8* promoter and/or enhancers) to form triple helical structures that prevent transcription of *PARK8* in target cells in the body. See generally, Helene, C., *Anticancer Drug Des.*, 6(6):569-84 (1991); Helene, C., *et al. Ann, N.Y. Acad. Sci.*, 660:27-36 (1992); and Maher, L. J., *Bioassays*, 14(12):807-15 (1992). Likewise, the antisense constructs described herein, by antagonizing the normal biological activity of one of the *PARK8* proteins, can be used in the manipulation of tissue, *e.g.* tissue differentiation, both *in vivo* and *for ex vivo* tissue cultures. Furthermore, the anti-sense techniques (*e.g.* microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a *PARK8* mRNA or gene sequence) can be used to investigate the role of *PARK8* in developmental events, as well as the normal cellular function of *PARK8* in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

In yet another embodiment of the invention, other *PARK8* therapeutic agents as described herein can also be used in the treatment or prevention of Parkinson's disease. The therapeutic agents can be delivered in a composition, as described above, or by themselves. They can be administered systemically, or can be targeted to a particular tissue. The therapeutic agents can be produced by a variety of means, including chemical synthesis; recombinant production; *in vivo* production (*e.g.*, a transgenic animal, such as U.S. Pat. No. 4,873,316 to Meade *et al.*), for example, and can be isolated using standard means such as those described herein.

A combination of any of the above methods of treatment (*e.g.*, administration of non-mutant *PARK8* polypeptide in conjunction with antisense therapy targeting mutant *PARK8* mRNA; administration of a first variant encoded by *PARK8* in conjunction with antisense therapy targeting a second encoded by *PARK8*), can also be used.

The invention will be further described by the following non-limiting examples. The teachings of all publications cited herein are incorporated herein by reference in their entirety.

EXEMPLIFICATION

Identification of the PARK8 locus with linkage to Parkinson's disease

A genome wide scan was performed on 117 patients and 168 relatives within 51 families using 781 microsatellite markers. Sixty-three percent (63%) of the patients were men, and eighty-one (81%) of the 117 had age-at-onset symptoms over 50 years, with a mean age-at-onset of 65.8 years. The comprehensive genome wide database has been established at deCODE genetics, Inc.

Encrypted patient identifiers, approved by the National Bioethics Committee of Iceland and the Data Protection Commission of Iceland (DPC), were used. Patients were identified as described in "Familial Aggregation of Parkinson's disease in Iceland", by S. Sveinbjornsdottir, A.A. Hicks, T. Jonsson, *et al.*, *N. Engl. J. Med.*, 343:1765-70 (2000). Briefly, patients identified through a total population survey carried out in Iceland in 1953-1963 were combined with an ongoing population-based study which identifies patients from a variety of sources including the Parkinson's Disease Society, neurologists, general practitioners and records of levodopa and other Parkinson's disease drug prescriptions, (Gudmundsson, K.R., *Acta. Neurol. Scand.*, 43, Supplement; 33:1-61 (1967)). All nursing homes and homes for the elderly in Reykjavik and those in approximately 70% of the rest of Iceland were visited to examine patients. Living patients were reevaluated and all relevant medical records were reviewed by two independent neurologists. Patients were considered to have Parkinson's disease if they have at least two of the following cardinal signs: tremor; rigidity, bradykinesia or postural instability while

other causes of parkinsonism were excluded (Hoen, M.M.and M.D. Yahr, *Neurology*, 17:427-442 (1967)). 951 cases of Parkinson's disease disease were identified which represents over 90% of the total number of patients diagnosed in Iceland during the last 50 years. Only patients who were related to other patients within 6 meiotic events (6 meioses separate second-cousins) as determined using deCODE's genealogy database were included. 117 patients and 168 of their first degree unaffected relatives were included. Informed consent was obtained from all patients and their relatives whose blood and DNA samples were used in the linkage scan. All personal identifiers associated with medical information and blood samples were encrypted by the DPC as described in "Protection of Privacy by Third-party Encryption in Genetic Research in Iceland," by J.R. Gulcher, K. Krisjansson, H. Gudbjartsson and K. Stefansson, *Eur. J. Hum. Genet.*, 8:739-742 (2000). The genealogy database was encrypted by the DPC in the same manner.

Genome-wide linkage scan

A genome-wide scan was performed using a framework map of 781 microsatellite markers. The marker order and positions for the framework mapping set were obtained from the Marshfield genetic map (*http://research.marshfieldclinic.org/genetics*), except for a three-marker putative inversion on chromosome 8 (Gulcher, J.R., *et al.*, *Eur. J. Hum. Genet.*, 8:739-42 (2000), and Jonsdottir, G.M., *et al.*, *Am. J. Hum. Genet.*, 67 (Suppl. 2), 332, (2000)).

Statistical significance was determined by applying affecteds-only allele-sharing methods (which does not specify any particular inheritance model) implemented in the ALLEGRO (deCODE genetics, Inc.) program. ALLEGRO is a linkage program developed at deCODE genetics which calculates lod scores based on multipoint calculations and is available for non-commercial use. Baseline linkage analysis uses (1) the S_{pairs} scoring function, (Kruglyak, L., *et al.*, *Am J. Hum. Genet.* 58: 1347-63 (1996)), (2) the exponentially allele-sharing model, (Kong, A., and N.J. Cox, *Am. J. Hum. Genet.* 61: 1179-88 (1997)), and (3) a family weighting scheme which is halfway, on the log scale, between weighting each affected pair equally and weighting each family pair equally (Yu, A., *et al.*, *Am. J. Hum. Genet.*,

67 (Suppl. 2) 10 (2000), Kruglyak, L., *et al.*, *Am. J. Hum. Genet.*, 58:1347-63 (1996), Gudbjartsson, D.F., *et al.*, *Nat. Genet.*, 25:12 (2000)). In the analysis all genotyped individuals who are not affected were treated as "unknown". P values were computed two different ways and the larger (less significant) one was reported. The first P value, which is computed based on the large sample theory; Z_{lr} = (square root of) (2 $\log_e$ (10) lod), is approximately distributed as a standard normal distribution under the null hypothesis of no linkage (Kong, A. and N.J. Cox, *ibid.*). Apart from the P value completed using normal approximation, because of the concern with a small sample behavior using ALLEGRO, a second P value was computed by comparing the observed lod score to its complete data sampling distribution under the null hypothesis (Kruglyak, L., *et al.*, *ibid.*). When a data set consists of more than a handful of families, these two P values tend to be similar. Conservatively, the larger of the two was reported. The P values can be adjusted for multiple comparisons. An adjusted P value of 2×10^{-5} corresponds approximately to a P value of 0.05 after adjusting for a genome-wide scan.

Genotypes of polymorphic markers provide uncomplete information about DNA sharing among patients. The degree of completeness can be increased by typing additional markers. This measure is closely related to a classical measure (Dempster, N.M., *et al.*, *J. R. Statist. Soc. B.*, 39:1 (1977)). Information equals zero if the marker genotypes are completely uninformative and equals one if the genotypes determine the exact amount of allele sharing by descent among the affected relatives.

Genetic and Physical mapping

High resolution genetic mapping and basepair resolution physical mapping were used to locate microsatellite markers and genes. Additional microsatellite markers were designed using genomic sequences from the January 9, 2000 data freeze assembly of the University of California at Santa Cruz (*http://genome.ucsc.edu*). Genotype data from 112 Icelandic nuclear families containing 812 individuals (sibships with their parents, containing two to seven

siblings were used to create a genetic map over the interval of *PARK8*. Misassembly or misordering of the contigs revealed were corrected and confirmed with selective coincident hybridization experiments of the BAC library. The combined genetic-physical map determined a most likely order of markers and the inter-marker distances have a resolution of about 0.5 centimorgans.

Parametric modeling

After obtaining a significant allele-sharing lod score, parametric models are fitted to the data to explore the contribution of the susceptibility gene to the disease population. When fitting parametric models, affected-only analysis was performed. As a result, only the ratio of penetrances are relevant. A range of single locus dominant, additive and multiplicative models was fit. With a dominant model, the penetrance of homozygous carrier of the at-risk allele is assumed the same as for a heterozygous carrier. The penetrance of carriers can be smaller than one and non-carriers are allowed to have non-zero penetrance. It is only the ratio of the penetrances that affect the lod score. An additive model assumes that the difference between the penetrances of a heterozygous carrier of the at-risk allele versus a non-carrier is the same as the difference between a homozygous carrier and a heterozygous carrier. With a multiplicative model, the ratio of the penetrance of a heterozygous carrier versus a non-carrier is assumed to be the same as the ratio of the penetrances of homozygous and heterozygous carriers. Compared to a dominant model, a multiplicative model tends to fit better with a higher assumed at-risk allele frequency and a lower penetrance ratio between the heterozygous carrier and a non-carrier. The additive model tends to be in-between the dominant and multiplicative models. With affected-only analysis, while one model may fit slightly better than another, it is usually not enough to be confident whether the true model of inheritance is dominant, multiplicative or recessive. With complex disease, none of these simple models are likely to be exactly true. The effect of a gene and its variants, it is believed, can only be reliably determined after the at-risk variant, or variants, are identified. However, apart from the statistical significance, the fitted parametric models provide an idea as to how much the gene is contributing to the

disease. For example, the corresponding contribution to the sibling recurrence risk ratio and the population attributed risk can be calculated. The sibling recurrence risk ratio is a measure of the magnitude of inheritance and is related to the power of detecting linkage. Population attributed risk is defined as the fraction of cases reduced if the population were to consist entirely of non-carriers of the at-risk allele. The best fitting dominant, additive and multiplicative models tend to give similar sibling recurrence risk ratios, but the corresponding population attributed risks are more variable.

Results of linkage and localization

The linkage scan was conducted on a total of 117 Parkinson's disease patients within 51 families. FIG. 1 is an example of the extended families used in the study, which include multiple Parkinson's disease patients. The patients and 168 of their relatives not diagnosed with Parkinson's disease were genotyped using 781 microsatellite makers. The data was analyzed and significance determined by applying affected-only, allele sharing methods. The methods do not formally specify a particular inheritance model, but instead search for genomic regions shared by affected relative by descent more often than expected under the baseline of no linkage. Lod scores based on multipoint calculations, which used the information from all the markers simultaneously were computed using the ALLEGRO program (deCODE genetics, Inc.).

Sex indicators have been shuffled for some individuals in the top two generations, and unaffected siblings, offspring and mates of the patients are not shown, to protect privacy. The solid squares and circles represent affected men and women, respectively. The patients in this family had ages at onset of symptoms between 66 and 78 years old. The slashed symbols represent deceased individuals for whom we have no DNA. Haplotypes shown are for 10 markers from a shared region spanning 37 markers, plus one marker just outside the region on either side. The shared region is represented by a solid box.

The allele sharing lod scores and the corresponding Z_{lr} scores for the entire genome using the markers in the framework map were determined. The most

prominent linkage was found to chromosome 1p32. Two lod score peaks of 3.9, separated by about 13 centimorgans, were observed near D1S2652 and D1S2846, respectively. Additional lod scores of 1.6 on chromosome 5 near D5S666, 1.2 on chromosome 7 near D7S661 and 1.1 on chromosome X near DXS8080 were also observed. While the lod score of 3.9 (Z_{lr}= 4.2, P=1.1 x 10^{-5} unadjusted on chromosome 1) was already genome-wide significant according to the Lander-Kruglyak criterion (single test P value of less than 2 x 10^{-5} which corresponds to a genome-wide adjusted P value of 0.05; Lander E., *et al.*, *Nat. Genet.*, 11:241-247 (1995)), the information content was below 60%, indicating that the marker data was still far from capturing all the allele-sharing information inherent in the material. It is more meaningful to have a lod score associated with high information content. Apart from ensuring that the results are a true reflection of the information contained in the material, it reduces the potential problem of multiple comparisons, as multipoint lod scores can increase or decease substantially as markers are added. The information content for linkage at the 1p32 region was increased by genotyping an additional 44 markers over an approximately 44 cM segment which spans the region between markers D1S2884 and D1S198.

Correct marker order and good estimates of inter-marker distances are extremely important for multipoint linkage analysis (Halpern, J. and A.S. Whittemore, *Hum. Hered.*, 49:194-196 (1999) and Daw, E.W., *et al.*, *Genet. Epidemiol.*, 19: 366-380 (2000)). However, the microsatallite marker maps in the public domain have limited accuracy at inter-marker distances less than 2 to 3 centimorgans. High-resolution genetic and physical mapping were used to increase the accuracy of the order of additional markers and to provide reliable inter-marker distances. With the additional markers and the map described herein, the lod score peak near marker S1S2846 dropped, while the lod score peak near marker D1S2652 increased to 4.9 with a corresponding Z_{lr} score of 4.8 (as shown in FIG. 2). This corresponds to a P value of 1.0 x 10^{-6} before adjusting for multiple comparisons, and a P value smaller than 0.005 after adjusting for a genome-wide search. The information content over the 20cM interval centered on the peak exceeds 0.85 and averages 0.95. The locus was designated *PARK8*. The peak centered near marker

D1S231, with markers D1S2874 to D1S475, telomeric and centromeric, respectively, defining a drop of one in lod score from the peak. The segment is estimated to be approximately 7.6 centimorgans in genetic length and approximately 9.5 million bases.

To understand the contribution of this locus to the population, a variety of parametric modes were fitted to the data. By specifying allele frequencies and penetrances that maximize the lod score, a lod score between 5.3 and 5.6 can be obtained for dominant, additive or multiplicative models. While this indicated that the current data does not allow for elucidation of a direct mode of inheritance, an estimate of how much this susceptibility gene contributes to familial risk can be obtained. With an additive model that assumes that the carriers of one and two at-risk alleles have respectively 30 times and 59 times the risk of getting the disease compared to the non-carriers, log scores between 5.5 and 5.6 are obtained for the at-risk allele frequency ranging from 0.4% to 2.8%. These models correspond to a sibling recurrence risk ratio ranging from 3.2 to 4.6. The best fitting dominant and multiplicative models give similar numbers, which is not surprising as the power to detect linkage with affected-only analysis is more directly tied to the recurrance risk ratio instead of the exact mode of inheritance (Risch, N., *Pairs Genet. Epidemiol.*, 1: 109-22J and *Hum. Genet.*, 46: 229-241 (1990)). Compared to the observed sibling risk ratio of 6.7 earlier reported (Sveinbjornsdottir, S., *et al.*, *N. Engl. J. Med.*, 343: 1765-1770 (2000)), the gene at this locus alone can account for a substantial fraction of the familial aggregation of late-onset Parkinson's disease. The difference can be attributed to other contributing genes and the shared environment of the siblings.

Identification of the RNF gene

The identified gene is RNF 11 (Seki, N., *et al.*, *Biochem. and Biophys. Acta.*, 1489:421-427 (1999) and encodes, in three exons, ring finger protein 11. This protein contains a modified RING-H2 domain (also called C3H2C3) similar to the RING domain (C3H2C4) found in Parkin, and which is characteristic of a protein with a role in the ubiquitination pathway. This pathway, important for the breakdown of protein in the cell, comprises a number of enzyme activities

commonly referred to as E1, E2 and E3 activity. E1 and E2 are ubiquitin conjugating enzymes while E3 is ubiquitin ligase activity, important for substrate recognition. Parkin, the mutated protein in juvenile and early onset Parkinson's disease, has been shown to function as an E3 ligase. By homology, RNF 11 may function in the important process of protein turnover. In a disease whose pathology is characterized by protein aggregates, the possible significance of RNF 11 and disturbances in its function should not be underestimated.

Upon sequencing of DNA samples from individuals in Iceland, 7 new SNPs were found in the RNF 11 promoter region. These were found at nucleotides 1196, 1393, 2064, 2218, 2435, 2599 and 2703 of SEQ ID NO: 1. One additional SNP (at nucleotide 3181) was found in the sequence data of Exon 1, five bases upstream of the start ATG codon and may affect the KOZAK sequence. Three exons were found which encode RNF 11. Exon 1 is located at positions 3059 to 3308, Exon 2 is located at positions 36385 to 36554, and Exon 3 is located at positions 37580 to 39685. All numbering is relative to SEQ ID NO: 1. The amino acids coding RNF 11 are represented by SEQ ID NO: 2. These SNP's have not been documented prior to this disclosure.

Transcription binding sites were identified by sequence homology to known transcriptional binding sites, such as those for BRN-2 (a neuronal-specific transcriptional regulator (nucleotides 1182-1197); MEF-2 nucleotides (2416 to 2437); GC or SP1 (nucleotides 2665-2677); ATF or CREB (nucleotides 2589-2600 and 2694-2705); NGFIC or EGR1, 2, or 3 (nucleotides 2779-2789) and HEN1 (nucleotides 2803-2824). SNP's were identified within the BRN-2 site, MEF-2 site, ATF/CREB (2589-2600) site, and ATF/CREB (2694-2705).

Western Blots

Western blots of patients' brain material who presented Parkinson's Disease (PD), Alzheimer's (ALZ), mixed (PD/ALZ) or control brain material with no disease, were run on a western using RNF-11 peptide-generated antibodies. The peptide sequence used to generate antibodies was KGVYDPGRDGSEKKIRE (SEQ ID NO: 3). It was observed that some patients contained significantly larger amounts

of RNF-11 and oligomeric forms of RNF-11 (higher molecular weights) than the control patients who did present the diseases. Further, patients with pathological features of both Parkinson's and Alzheimer's demonstrated RNF immunoreactivity with senile plaques as well as Lewy bodies and neurophil threads reminiscent of Lewy neurites.

Additionally, experiments were done with yeast two hybrid experiments were performed to identify specific interacting protein(s). Multiple libraries were screened and individual clones were isolated from different brain libraries. Specific interactions were shown with U33821.1, human TAX-1 binding protein (TXBP151); AF095745.1, human ubiquitin protein ligase (ITCH); AF110265.1, human EGF receptor substrate (EPS15R); AF241230.1, human TAK-1 binding protein 2 (TBP2); AF293385.1 human PLIC-2; AF301463.1, SMURF2 (ubiquitin ligase); AF062085.1 human epsin 2b; and U22897.1 nuclear domain 10 protein (ndp52).

While this invention has been particularly shown and described with reference to preferred embodiment thereof, it will be under stood by those skilled in the art that various changes in form and details may be made therein without departing form the spirit and scope of the invention as defined by the appended claims.

CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule comprising a *PARK8* gene, wherein the nucleic acid molecule comprises at least one single nucleotide polymorphism selected from the group consisting of: G at nucleotide 1196, C at nucleotide 1393, T at nucleotide 2064, T at nucleotide 2218, C at nucleotide 2435, C at nucleotide 2599, G at nucleotide 2703 and G at nucleotide 3181, and combinations thereof, the nucleotide positions being relative to SEQ ID NO: 1.

2. The isolated nucleic acid molecule of Claim 1, wherein the Parkinson's disease gene has the nucleotide sequence of SEQ ID NO:1.

3. An isolated nucleic acid molecule comprising a nucleotide sequence having all or a portion of the promoter region for the gene encoding ring finger 11 protein and, comprising at least one single nucleotide polymorphism selected from the group consisting of: G at nucleotide 1196, C at nucleotide 1393, T at nucleotide 2064, T at nucleotide 2218, C at nucleotide 2435, C at nucleotide 2599, G at nucleotide 2703 and G at nucleotide 3181, and combinations thereof, the nucleotide positions being relative to SEQ ID NO: 1.

4. A method for assaying the presence of a first nucleic acid molecule in a sample, comprising contacting said sample with a second nucleic acid molecule comprising the nucleic acid molecule of Claim 1.

5. A vector comprising an isolated nucleic acid molecule corresponding to exon 1, exon 2, exon 3 or combinations thereof of the *PARK8* gene,

operatively linked to a promoter for the *PARK8* gene having at least one single nucleotide polymorphism selected from the group consisting of: G at nucleotide 1196, C at nucleotide 1393, T at nucleotide 2064, T at nucleotide 2218, C at nucleotide 2435, C at nucleotide 2599, G at nucleotide 2703 and G at nucleotide 3181, and combinations thereof, the nucleotide positions being relative to SEQ ID NO: 1.

6. A recombinant host cell comprising the vector of Claim 5.

7. A method of diagnosing a susceptibility to Parkinson's disease in an individual, comprising detecting a polymorphism in a Parkinson's disease gene, wherein the presence of the polymorphism in the gene is indicative of a susceptibility to Parkinson's disease.

8. The method of Claim 7, wherein the polymorphism of the gene is selected from the group consisting of nucleotide position 1196, 1393, 2064, 2218, 2435, 2599, 2709, 3181 and combinations thereof; wherein the nucleotide position is relative to SEQ ID NO: 1.

9. A method of diagnosing a susceptibility to Parkinson's disease, comprising detecting an alteration in the expression or composition of a polypeptide encoded by *PARK8* gene in a test sample, in comparison with the expression or composition of a polypeptide encoded by *PARK8* gene in a control sample, wherein the presence of an alteration in expression or composition of the polypeptide in the test sample is indicative of a susceptibility to Parkinson's disease.

10. The method of Claim 9 wherein the alteration in the expression or composition of a polypeptide encoded by *PARK8* gene comprises expression of a polypeptide in a test sample that differs from a variant polypeptide expressed in a control sample.

11. The method of Claim 9 wherein the alteration in expression results in either a higher amount or lower amount of polypeptide compared to the amount of polypeptide expression in the control sample.

12. A method of identifying an agent which alters activity of a polypeptide encoded by a *PARK8* gene, comprising:

a) contacting the polypeptide or a derivative or fragment thereof, with an agent to be tested;

b) assessing the level of activity of the polypeptide or derivative or fragment thereof; and

c) comparing the level of activity with a level of activity of the polypeptide or active derivative or fragment thereof in the absence of the agent,

wherein if the level of activity of the polypeptide or derivative or fragment thereof in the presence of the agent differs, by an amount that is statistically significant, from the level in the absence of the agent, then the agent is an agent that alters activity of the polypeptide.

13. An agent which alters activity of a polypeptide encoded by *PARK8* gene, identifiable according to the method of Claim 12.

14. An agent which alters activity of a polypeptide encoded by *PARK8* gene, wherein the agent is selected from the group consisting of: a Parkinson's disease receptor; a Parkinson's disease binding agent; a peptidomimetic; a fusion protein; a prodrug; an antibody; and a ribozyme.

15. A method of altering activity of a polypeptide encoded by *PARK8* gene, comprising contacting the polypeptide with an agent of Claim 14.

16. A method of identifying an agent which alters interaction of the polypeptide encoded by *PARK8* gene with a Parkinson's disease binding agent, comprising:

a) contacting the polypeptide or a derivative or fragment thereof, and the binding agent, with an agent to be tested;

b) assessing the interaction of the polypeptide or derivative or fragment thereof with the binding agent; and

c) comparing the level of interaction with a level of interaction of the polypeptide or derivative or fragment thereof with the binding agent in the absence of the agent,

wherein if the level of interaction of the polypeptide or derivative or fragment thereof in the presence of the agent differs, by an amount that is statistically significant, from the level of interaction in the absence of the agent, then the agent is an agent that alters interaction of the polypeptide with the binding agent.

17. A method of identifying an agent which alters expression of a *PARK8* gene, comprising:

a) contacting a solution containing a nucleic acid of Claim 1 or a derivative or fragment thereof, with an agent to be tested;

b) assessing the level of expression of the nucleic acid, derivative or fragment; and

c) comparing the level of expression with a level of expression of the nucleic acid, derivative or fragment in the absence of the agent,

wherein if the level of expression of the nucleotide, derivative or fragment in the presence of the agent differs, by an amount that is statistically significant, from the expression in the absence of the agent, then the agent is an agent that alters expression of *PARK8* gene.

18. A method of identifying an agent which alters expression of *PARK8* gene, comprising:

a) contacting a solution containing a nucleic acid comprising the promoter region of a *PARK8* gene operably linked to a reporter gene, with an agent to be tested;

b) assessing the level of expression of the reporter gene; and

c) comparing the level of expression with a level of expression of the reporter gene in the absence of the agent,

wherein if the level of expression of the reporter gene in the presence of the agent differs, by an amount that is statistically significant, from the level of expression in the absence of the agent, then the agent is an agent that alters expression of a *PARK8* gene.

19. A method of identifying an agent which alters expression of a *PARK8* gene, comprising:

a) contacting a solution containing a nucleic acid of Claim 1 or a derivative or fragment thereof with an agent to be tested;

b) assessing expression of the nucleic acid, derivative or fragment; and

c) comparing expression with expression of the nucleic acid, derivative or fragment in the absence of the agent, wherein if expression of the nucleotide, derivative or fragment in the presence of the agent differs, by an amount that is statistically significant, from the expression in the absence of the agent, then the agent is an agent that alters expression of *PARK8* gene.

20. The method of Claim 19, wherein the expression of the nucleotide, derivative or fragment in the presence of the agent comprises expression of one or more variant(s) that differ in kind or in quantity from the expression of one or more variant(s) in the absence of the agent.

21. A method of altering expression of *PARK8* gene, comprising contacting a cell containing *PARK8* gene with an agent of Claim 20 identifiable according to the method of Claim 19.

22. A method of identifying a polypeptide which interacts with a *PARK8* polypeptide, comprising employing a yeast two hybrid system using a first vector which comprises a nucleic acid encoding a DNA binding domain and a *PARK8* polypeptide, variant, or fragment or derivative thereof, and a second vector which comprises a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a test polypeptide, wherein if transcriptional activation occurs in the yeast two hybrid system, the test polypeptide is a polypeptide which interacts with a Parkinson's disease polypeptide.

23. A Parkinson's disease therapeutic agent selected from the group consisting of: a Parkinson's disease gene or fragment or derivative thereof; a polypeptide encoded by a Parkinson's disease gene; a Parkinson's disease receptor; a Parkinson's disease binding agent; a peptidomimetic; a fusion protein; a prodrug; an antibody; an agent that alters Parkinson's disease gene expression; an agent that alters activity of a polypeptide encoded by Parkinson's disease gene; an agent that alters posttranscriptional processing of a polypeptide encoded by a Parkinson's disease gene; an agent that alters interaction of a Parkinson's disease polypeptide with a Parkinson's disease binding agent; an agent that alters transcription of variants encoded by a Parkinson's disease gene; and a ribozyme.

24. A pharmaceutical composition comprising a Parkinson's disease therapeutic agent of Claim 23.

25. The pharmaceutical composition of Claim 24, wherein the Parkinson's disease therapeutic agent is an isolated nucleic acid molecule comprising a Parkinson's disease gene or fragment or derivative thereof.

26. The pharmaceutical composition of Claim 24, wherein the Parkinson's disease therapeutic agent is a polypeptide selected from the group consisting of SEQ ID NO: 2.

27. A method of treating Parkinson's disease in an individual, comprising administering a Parkinson's disease therapeutic agent to the individual, in a therapeutically effective amount.

28. Use of a Parkinson's disease therapeutic agent for the manufacture of a medicament for the treatment of Parkinson's disease.

29. A Parkinson's disease therapeutic agent for treatment of Parkinson's disease.

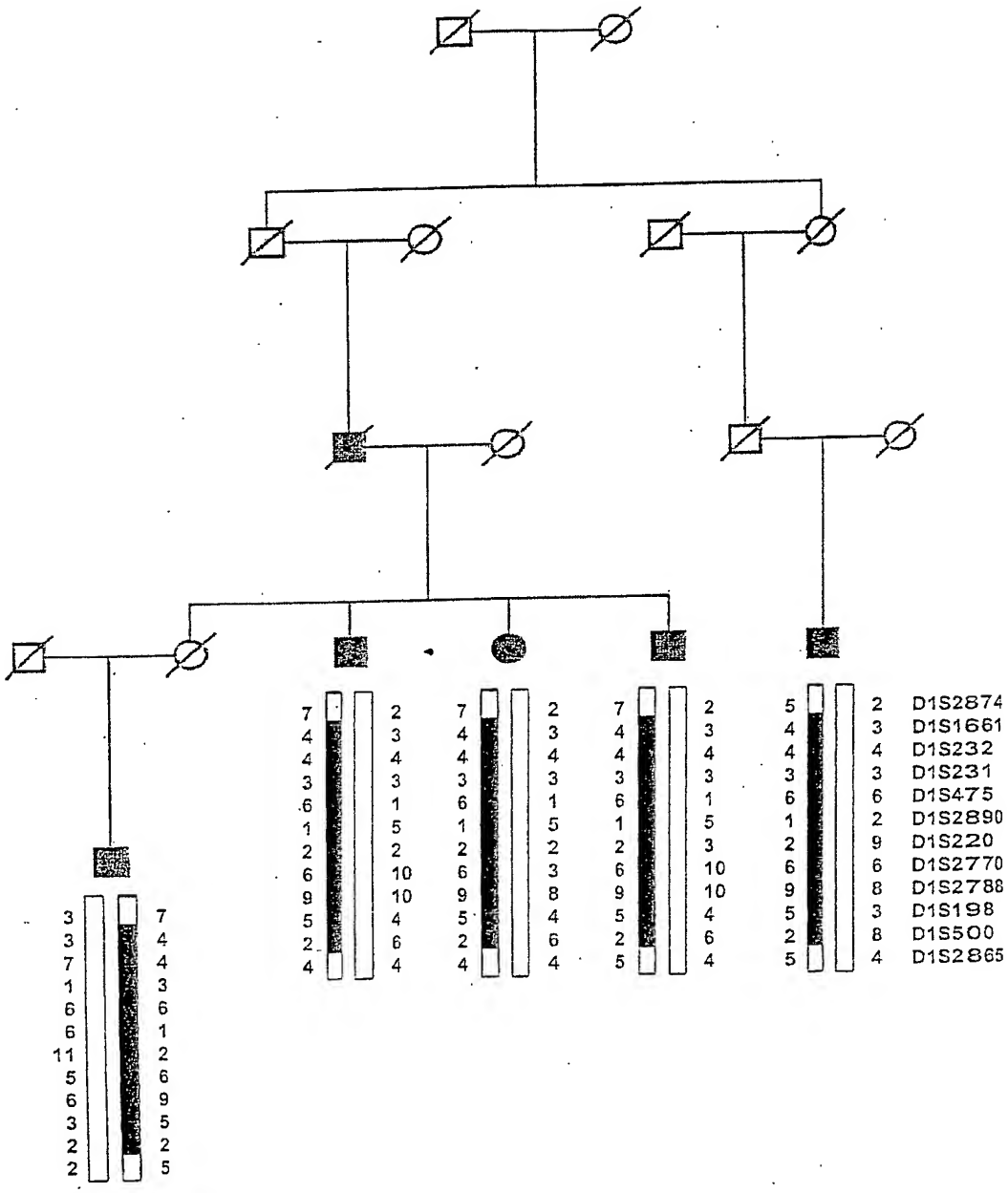
D1S2874
D1S1661
D1S232
D1S231
D1S475
D1S2890
D1S220
D1S2770
D1S2788
D1S198
D1S500
D1S2865

FIG. 1

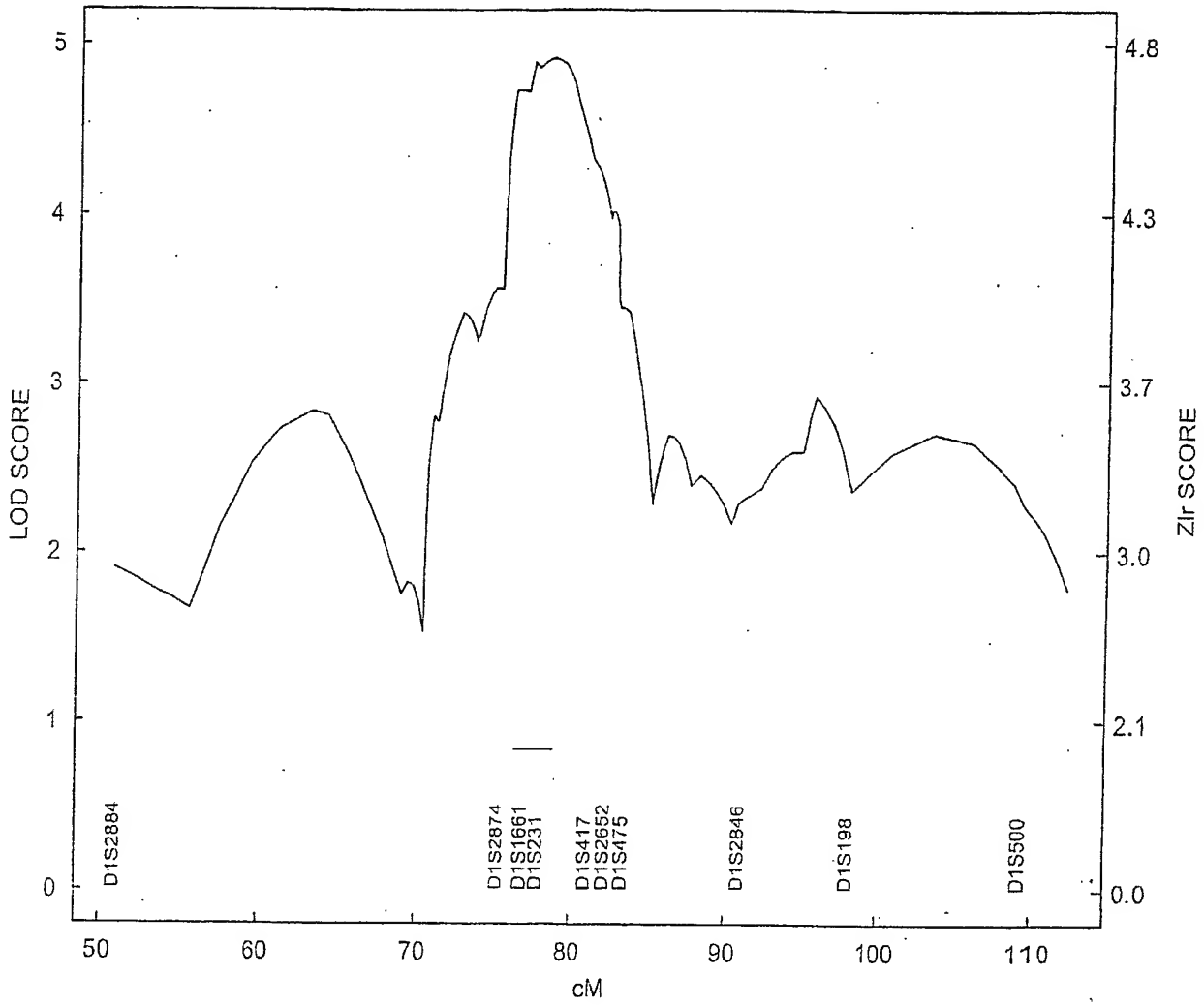
LOD SCORE
Zlr SCORE
cM
0
1
2
3
4
5
0.0
2.1
3.0
3.7
4.3
4.8
50
60
70
80
90
100
110
D1S2884
D1S2874
D1S1661
D1S231
D1S417
D1S2652
D1S475
D1S2846
D1S198
D1S500

FIG. 2

FIG. 3A

Nucleic Acid Sequence of RNF-11 gene

```
cccagctatt caggaggctg aggcaggaga attgcttgaa ccggggaagt agaggttgcg 60
gtaagccaag atcgcaccat tgcactctag cctgggcaac acaagcaaaa ctctgtctca 120
aaaaataaaa attaaaaaaa attaaaacta actaaataaa tctatgtctg gccaggcata 180
cagtggctca cacctgtaat tccagcactt tgggaggctg aggcgggagg atcgcttgcg 240
cccaggagtt tgagaccagc ctgggcaaca tggcaagact tcgtctctac caaaaacaaa 300
aacaaatgca aggcatggtg gtgtgcatct gtagttctag ttactggaga ggcctgaagc 360
aggaggatca cctgagccca ggaggttgaa gctgcagtga gcctcgatca agccagcctg 420
ggtgacagag tgagaccctg tctccaaaac aaacaaacaa tgtctacaaa ttctagcact 480
gtgtaaatat tatgtcaaat aattgtataa attattatta ataaaaattt ctttaggttt 540
gacaataatt tgtggggtaa attttctcac tttgtaaaaa ttccagggta cctctgattt 600
ctggtaaatt tagccaatca taaattaaat gagttcttgg aagctcaaca attccaaaac 660
caaaattgtg agtcagtttg ttctttcttt ccttctttcc tttctttttc ttttttctttc 720
tttctttctt tctttctttt ctttctttct ttcttttctt tctcttttct ctttctcttt 780
ctctcccttc cttctttcct tccttcctgc ctcccttctc tctctctctc tttctccttc 840
ctttttttt tgagactgag tcttactcac tccctcgccc acgctagagt gcagtggcac 900
aatctgggct cactgcgctg caacttccac ctcccaggtt caagcggttt tcctgcctca 960
ccctcccgag tagctgggac tacaggcgtg caccagcaca cctggctagt ttttgtattt 1020
ttagtagaga tggggtttca ccatgttggc caggctactc tcgaatccct gacctcaagt 1080
gatccacccg ccttgacctc ccaaattgct gggattacag acgtaagcca ccacgcctgg 1140
cccaaaatta tgagtctttt caaaaatttt atatcaaaaa ggtgatattt aatgtggaat 1200
ttgagtgcat ttaaataaga tagtgacttc gtagaatagg gactgcagag taagagtgtt 1260
cagacccatt cgagtttaat atagccctac ctctaagaat cctttctgtt cctcactcaa 1320
atgaactaat tgctgtcttt cagttcctct aaaaaccttg gggtccgtgg cagggtgtgg 1380
gcttaggact cacctgatta cttgtcttcc caggtggtct ctgggctcta tcaggaacaa 1440
gccaggtatg atcatccttt tccccagtcc ttagcgtggg gtctggtata cacacgcaca 1500
cacaacttca gtaaatggtg aatgaaacga aaggcagtca agatccacct gttgaatgaa 1560
cgtctatcag gagcactgtc atccttcttg cttccttctt ttgcatatgt tgtttgagct 1620
gctcattcaa ccaaggagag agcatactct accttgggca gtgtgccagc cttgcacaga 1680
gggacagatg attcagcttg atcaggctgt gacagagaaa agctcagggt ctgagagcac 1740
acaggatgcg actaacccag actggagggt caggagagtc tccccagggg aaatgaagtt 1800
gcatccagaa ggattcctga gaataatgat tagcctgatg agaagcagaa aggttgggag 1860
tggagggaag tgggtttctg tgagagaaac aatatggaca aaggccccag gcagaggcaa 1920
tagtattcca gataacgaaa agcattttag agagtctagt ttagaatttg tgaaggaggt 1980
gacaatgctg ctaatgagat tgaaagggag tcagatttgt aaaccattta aggattgtct 2040
taaggacagc cttgaggaga cactgaaggt tttaacagga gaatgacatg atccgattta 2100
aattttaaaa agataactct gtcaactgtg ccaacagtga attggatgtg ggcaatgatg 2160
aactcaggaa cagctgagag tctattgcag aagccaaggt gagagatgat ggtggcttgg 2220
actttggtgc tgggactgga aagaaatgaa caatctgata tacaactgga agaataatct 2280
ctctggctgt gatgtggaag gataaggacg tgattggaca ggcagggatg gttcagtaac 2340
gagctaggac tttattgtga agggataaag gggaagcttt taacatttta atcatgggac 2400
tgatccattc cgatatgggt ttatataaat agaccgcttg gcggagggtg gtgagaagac 2460
tgtcaggcag tgcacaggta agaactgagc cctagacggg gcaggagcaa tgaatgtggg 2520
gaggcgcggg agacgggaga ttggtttcgg agccaggtac ggagccaagc agtcagagag 2580
aagcaaaaat tgacgcagca ccgagaacag ctgctagtgg agaggggat gggcgccgcc 2640
ggagccgccc tcgcctcctg gcgccgggggg cgggaccaga gcacgcacgg cgcgatgacg 2700
cagtgcccga gccggcctcg ttctggctgc cgccgtctcg ctgaggcgga gtaggatgag 2760
gcccgcggcc gtggctccgg cgtcggcggg ggcagcagca gctgaggcag cagctgaggc 2820
agccgcgacg gccgcgcccc cccccgctga cctggattag gcccagcagc tccgtggccg 2880
ccgccgcccc gcgggggggc ggggtgggga agtgcttcgt ctccccgccc ccccccccc 2940
atcgctgcct cgcaggaggc gaacgccgcg accccagcgg agcccgcggc ccagccttga 3000
tccccaacc ccgggggctg gcatgagcgg cccctcggcg gcaccgtggg gcggtggagt 3060
cgcctccgcc tgatccccgg cctgtcgccc gaccccacct cgccaaccga ggcggaccgc 3120
ggagtgtgcg aacgacccca ccgctgcttt ctcctccccc agatcacgca ccccagctcc 3180
ggaagatggg gaactgcctc aaatccccca cctcggatga catctccctg cttcacgagt 3240
ctcagtccga ccgggctagc tttggcgagg ggacggagcc ggatcaggag ccgccgccgc 3300
catatcaggt aggggagggt gtgtgttggg gggagcgagt agaggcagct ctggcggaga 3360
```

FIG. 3B

```
ttgaggggggc ttcggggccg tctctgcccc tggcttcggc aaggcctggc ctgggcggca 3420
ttgacccctt agggctggat ctgagggcat ttgctctctg atctcagagt aaggttgatt 3480
cttcgcctgc cctcgggcac cgtgtgggga tgataccctc tgcccttgag ttctgggatc 3540
agagtatgtg atcacatccc cttggccagt gtctggtggt tcgcttgtgg agtttgggag 3600
tgacattttc ctgtctttat gtctgaatcg tatgatttcc tcatcgcacc gccaactgga 3660
ttatgggttc tggataggtg gttgccaagc accacccccc cgccccattt atgttttcat 3720
gtgtggtcgt cctcttaccc acctccctcg tctttttgtt aattggatct ggcaggtttt 3780
ccctctgctt gggagacgga tctccaagtg gtgattttaa ccctagcgac atcgagttac 3840
ttgtgtgtgt gtgtatatat atatatatgt gtatatatat atatatatat atgtatataa 3900
attttaatcc ctcctctttt tgctccaccc atatgacaca ccttttaaac gtgctgactt 3960
aagaaaactt gattagcttg aactttggaa acctttaaac ctggcttggg aaattccttc 4020
ggttactcgt gttctccctt tgtatcaaag actctcctaa atacgttgta ctgcttttg 4080
tacaataatc agtagtccca gatgtgttgc agaattcgag tgaagacctc ttaaactcaa 4140
aataggtatg tggatgttat aatccacaat agatgtgtga aaacctttcc tgtacttttt 4200
ccattttaga tgatattgaa tacacccttt tccagagggc ttcaatagta gtagtgctta 4260
atttcttcct gaagagttta gttttctgac tgggtaggct cctttgttgt ttcatgtgat 4320
gacatcatta tctagtgttt tcttttctca ttcctgtgta aaccaatgtc tcatcacacc 4380
agaaaaacct gatcatatat atgctcgttt tctgaataaa attcttctgg gggcatttca 4440
tatataccag taaggtccag tcttgattgt ggtattacag aagcctccat gacttctccc 4500
tcttacgttc tatattccag gcctaccaaa tacaccaggc tgttcacagc tttgttttgc 4560
tccagttcct ctctgcttgg agcccttcgc ccgggtcagc caggtggaca tctaatcctc 4620
tttgagaact cagctttcat gtctgaagtc ccagcattcc cttccttccc agcccgaatc 4680
ccttgtgtcc ccactgaaac ctctacactt ccataagcaa tatctataac actttagtgt 4740
tgtatctttg ttagatgtgt gtctcatcct tgtaatcgcc tccttcagtg caagctgcgt 4800
gaaggtagag aatctgtgtt tccggtttgc tgttgtatcc ctcgcatctg cctcagtagg 4860
gactcagtaa ctgggagata aattacgagt aaaagccctt gagaacaggg actctccct 4920
cgctcagttt ctggcacata agcacctaaa tgctgaatta attaatgaag ttgaatttag 4980
tattttctgt atttacacaa atatttttc caaagtaagt tgtcacaggg tagtctctta 5040
aaaatcaaag ctgaatctgg gtgtctttac aagtaccttt gagtgaagca agcaagctat 5100
gtttatcctt cactgtcttt ccctcctatg tttactatac aacaaaagtg cttatttaac 5160
ttttttttt taatgagatg gagtctcgct ctgtcacccc ggctggagtg cagtggcaca 5220
atctcactgc ttactgcaac ctccacctcc cgggttctag cgattctcct gcctcagcct 5280
cctgagtagc tgggattaca ggcgcgtgcc accacaccta gctaattttt gtatttttag 5340
taaagacggg gtttcaccat gttggtcagg ctggtcttga actcctgacc tcaagcagtc 5400
cacccacctc agtctcccaa agtgctagga ttacaggcat gagccaccgc atctggcatg 5460
aagtttttat tttaaaatct cttatgctcc ctaccctcac ccccttctat ttttaaaaaa 5520
atagagacag ggactcctat gttgcctagg ctggtcttga actcctggcc tcaagcaaat 5580
tctcctgcct cctatccccc ttcttaattg aatatctttg acttgttctc tttctcctat 5640
ccttatgata attgtactat taaaggtact agaatttatg gaaactgtgt ttgccttttc 5700
cagaagaact aggtgatgat gagggatgag gggtttttgg tgtcatgaag gaacttttt 5760
tggaaatgtt atatttttga cagtgggagt tttttatggt tgcagtttcc tgaagtgaat 5820
gacgcagact aatatacaaa ttcagtatca atcagaagaa aattacttta aaacatggga 5880
gcataaaaag aaacttttgg agaaaaagta ttgtccccca cccccaaaga aataagatct 5940
gtcttaggaa tggttatttt agatgatctg agaaacagca atccaaggtt atttattcat 6000
ttatttgagt tgagaatctc cctctgtcat ccaggttaga ctggattgct cacttggcag 6060
cctcaaactc ctggcttcag gtgatcctcc tccctcagcc tcccaaagac tgggattaca 6120
ggcgttagct acctcacaca gcagatttat agtattctca gttgataaaa gtaacttttt 6180
caacaaatta tatgtaatta acagtgaata gctgattgta aaggaacatg ttctcttgag 6240
gtgagtagtt aaaagatgtc attgtttgct gctttatatg tacaatgtac atataatgac 6300
aagagccagt tacagttatt tcctattaac atctgacaaa tttgttttta tttatttttg 6360
ccttaaagta aatccctgca ccattatgat ccaagtttgt aggaagtgta tatttctttc 6420
catggtttct tttgtagcct cctaacagtt atccctgctt tcacccttgt cctcgctaag 6480
gcctgtcctc aatacaacag ccacagttat tttggtaaaa ctgagcatga cagtctcccc 6540
aaccctctgt gtggatcctt gtttcattca aagtaaaaat cagattcctt atagtgtcgc 6600
ataaacccta cataatctgt acccccttgtt tccacgcatg tcccttgaca cctacagcta 6660
acccctttgt tggcttccct ccagcttccc tgtcctcatt gctattctag gacctttgtc 6720
ctccttattc tggaatactc ttccccagat ttcatggctg actgactccc ttacctcctt 6780
caagtctttt ctcactgccc ccttaaacct tctctgacca ccctgtttga aagtgcaacc 6840
```

FIG. 3C

```
tgccctaata acctggaagt ccctctcacc ctgttgtttt ccccccccaat accatttatt 6900
acattccaac atgtgttaga aagttgcgtt gtctattacc tgtcttctcc cactaggtag 6960
aaaggaggaa aggggatttt ttttttaat tggtgtttcc taagtactta taacagtgcc 7020
tggcacatgt taggcattca gtaatttggt tgaacgaatg aatagatttc ttttaaatga 7080
actaatgtat ttgaggcagg tatagttgca ctgcttacaa gaactctgca tattcctagt 7140
gccatttttg cttctatatt gtagatcagg ttttttttttg tttgtttttt tgttttgtga 7200
ctgggcctcc ctgtgttacc caggctggtc tctctctgta tatatatgtt tgggtttttt 7260
tgttttctct tttttgagac caagtctcac cttgttgccc agctagagtg cagtggtgtc 7320
atcttggctc tgcctcttgg gttcaagcga ttctcctgcc tcagcctctg gagtagttgg 7380
tactacaggc atgcgccacc aaggccaact gatttttgta cttttggtag agatggagtt 7440
ttgccatgtt gaccaggctg gtctcaaact cctgacttca ggtgatctgc ccaccttggc 7500
cttccagagt gctgggatta caggcgtaag cccctgagcc tgccctttat attttttaaa 7560
tttaaaaaat agagacaggg tctcattatg ttgctagtct caaactcctg ggctcaaggg 7620
atactcctgc ctcagcctct tgagtagcta gggttacagg tgcacttcta ctctgcctgg 7680
ctcaggtttt aaaaaaaatt tatttccttt agagatgagg tcttgctgtg ttgcccaggc 7740
tggagggcag tagctattca caggtgcaat tatagtgcat tgcagcctcc acttcctggt 7800
ctcaaaggat cttgtggctt cagcctcctg agtagctggg actacagggt tacaccaccg 7860
tgtctagctg ttgtagatca gttttaagga atcctaaaaa gtcaaattcc agtaatatta 7920
aatacagcaa ttggtttaca ctaagtgtta agtgatttag gggtttttac cgtgtatata 7980
tatcaccgca agatatttat gtaatgggaa agtacttggg ggagaaaaag aaagtacttg 8040
aatccctatt cttgaagtat tggatgcaaa gagatttagg gtgtttgtca aattagaaag 8100
tagaaacaaa tttgtcgcac cctctaccca actcccttat tcctaaatgt ttaaaaataa 8160
tacagaggaa atctgcccag cgaaggcaat cttggaagga cagtgagtgg gaggataagc 8220
caggaaaggc tgaagaaggg aggtccatga gaggggagg ccaggtaggc agcccagccc 8280
tgaaaggcat atgtaaaaca gcttgaggta cccctgtgtc actgtcaaga gctttgttat 8340
gtgctaaatt gagggaggca ctttgttaaa agagtgtaat attcattgaa caacaaacca 8400
tgttacattt tctaatctga atatataaaa ataaggatag gtgtctctca attttgcagt 8460
agttatatag tttgagattc gccagtcttc ttccccgacc tatacttttt gattgataca 8520
tttgaaattt taaaatcctg agaaaataag tttgtatagc attgaatgca agtatagctc 8580
tgcttttaga gtcagattgt cctgtgttca aattctggct ctgccgctta catgtaaaat 8640
atgaaaaagg aaagaattta tcagattgtt ttggggagct tttttgacaa ataagaaaat 8700
gtaacatttt ggatatttca aattgtgtct gcataaagct ttcttcaaga tacctttttt 8760
ttttttttt tttggctggg cacagtggct cacacctgtc atcccagcac ttaggtggca 8820
gaggcaggag gatcacttga ccccaggaat tcgagaccag cctgggcaac aaagtgagac 8880
ccccatctct attaaaatac caaacaaaaa caaaaaacaa acaaaaccaa aaaccctctt 8940
ttgttgttaa aaggattggt ttatcctgta gaatgcacta aaaataaccc ccttccttat 9000
aaaaaatcta gaatgtactg cttatttta agcatgcaat gctataatgt tttgtggggg 9060
atgggcttgt ttttttaga gatggagtct cactttgccc aggctagttt cgaactcctg 9120
ggctcaagtg atcctcccac catgtctccc agtgttggga ttacggatgt gagtcactgt 9180
gccctgcctg taatggtata ttaatgcctc attttcctga taacttacga agttaaattt 9240
agataattaa agggttgtgt ttttttatta gcattttgtt taacatttgg gtggaaaata 9300
tctccaaatt tatttcacgt atctaggcct tattatacaa ttattcagct tttaacaatt 9360
taccttttg aagttttatc cctcccccttc tttatgctca agaagttcag ccttcttatt 9420
gaggggcaga tcaactaact ccacctttgt taataaagcc ttcactgacc attttttttcc 9480
tgtctctgct gatttgtcat gtacagtttt tttgtgtttg taacatttat ctggaaatca 9540
tatattgcct tattagcttt tgttgcttta acagttgttt ctccacatca gtttcctcaa 9600
gaatgagaat gatacattaa acttgtaaga tatcttagtg gagtagaatg ttaccttcca 9660
tattgtagac acccagtgaa tatttgatga tagactaaat tagaactgtg acttcttacc 9720
ttgtaaggtc accatcataa ttgtcctcat cagttcttac agcaacataa ttgatgtctc 9780
ctcctcattc attttatacc atgttgccag ataaatttca tcccggatca gatcacttgc 9840
ctcataatcc ttaatgattg ttattattat tattattatt ttttgaaaca gagtcttcac 9900
tctgttgccc aggctggagt gcagtggtgc cacctcagct cactgcagcc tccgcctccc 9960
gggtccgagc aattctcctg cctcagcctc ccgagtagct gggattacag gcatgcgcca 10020
ccacacccag ctaatttttg tatttttaat agaaatgggg tttcaccata ttgctcaggc 10080
tggtgtgcaa ctcctaacct caggtaatcc actcaccttg gcctccgtaa gagctgggat 10140
tacaggcgtg agccaccacg cccggccctt aatgattctt tattaggact aaccactgaa 10200
gatatctgac ttagcattca aagtcatgat ttggctctac atatactttt tcagctcttc 10260
atagccttgt gctttctttt atacaacctg tatactccag tcatactgag cagtttgctg 10320
```

FIG. 3D

```
tattccatgt atgtcttctg ctttcctgtc tttgcccatt ctgtatccat tgcccagaat 10380
gttcttcctt aatccttccc agtctcattt ctaatatcca gggcccagtt caaatgccaa 10440
gaatccacga attctgattt ctcctacata cacttttctt aatctgaact cctttattca 10500
tgcaccattc ttaagacact ttacccttct aatttttaatt atttgtgaca gtgtcacatc 10560
tgctctttta gaattcaagt ctggttcatt ttcacattgc tagaactaca attttcacaa 10620
agcacgcact tatactttag tgaatgtaat gatcagcatt ttgctgtaag agtgtgcaat 10680
tgtatggttg ttagtgtcta ttaagcattc ccagaatcct gtgctttagt gtttggccta 10740
tcacttctaa tttataaagt acgttgagct ttgggaaagt taatttcatt atatattttt 10800
tgaatttgtt tttttttttg tttgtttgtt tgagacgaag tctcgctgtg tcacccaggc 10860
tggagtgcag tggcgcgatc tcagctcact gcaagctccg cctcccgggt tcacgccatt 10920
ctcctgcctc agcctgtccc gagtagctgg gactacaggt gcccgccacc acacctggct 10980
aatttttttgt atttttagta gggacggggt ttcaccgtgg tcttgatctc ctgaccttgt 11040
gatcctcctg cctcggcctc ccaaagtgct gggattacaa gcgtgagcca ccgcgcccag 11100
cctgaatttg gttttgtgtt ccatgaagtt cagcttgcct aggtgacttt tataaataag 11160
taaccagtta tttgttattt gtgtgacccc attgttgcag gatttttctt tgtttctgcc 11220
tgacctctca ttctgaatgt gctttaataa aggctgattt tttggcatta agattgtatg 11280
catcactaag agtaagatag tctgttcaga gtttaattct agagctaatt atagggtaac 11340
ttgagaagat ggagatggga tggtgagtta cttggtacat aagatctgag cacttgattt 11400
tcttaaaata ttttttaggt aatagtagaa acctttgttg cagaattttg atctttaaaa 11460
gttttaatag aaatacttat gtagattggg attctttttc tttctttgag acataatctt 11520
actctgttgc ctaggctgaa tgtagtggca tggtcacggc tcactgaagc ctcaacctca 11580
cgggcccaag ccatcctccc acctcagcct cctgagtagc tgggactata ggcgtgcacc 11640
actttgcctg gctaattttt attttttgta gagacgggtc aagcagtcct cccatctcag 11700
tctcctgaat agctgggacc acaggtgcat accaccacac ccagctattt tttgtatttt 11760
tagtagagac ggggtctcac tatattgccc aggctggtct caaactcctg aactcaagca 11820
gtccacccac ctcggcctcc caaagtgctg agattacaag tgtgagccac catgcccacc 11880
caacagcctt ttttttgga gatggagtct tactccatcg cccaggctgg agtgcagtgg 11940
tgcagtctca gcccactgca acttctgcca cctgggttca agcgattctc atgcctcagc 12000
ctcctgagta gctgggatta gaggcaagca ccaccacacc tggctaattt tttgtatttt 12060
tagtagagat ggggtttcac catgttggcc agactggtct taactcctga cctcgtgatc 12120
cacccaccta ggcctaccaa agtgctgaga ttacaggcgt gagctactgg gctggacacc 12180
aacagcctat ttttaataac ccatatgagt acaggacaga aatacatccc taactctgta 12240
cctgaagtaa ggaaggtaat acataaaaga catatttgca aacgacttga ggcaagaaaa 12300
gaaatacaac acatctctaa ttctgattgt atctgcaatc aacctcttat cacttccctc 12360
ttcatgtgca gaatttacct tggctgagac acagtcaaga aagaagacag gtggccaggc 12420
acagtggctc acacctgtaa tctcagtact ttgggaggct gaggtgggcg ggtcacctga 12480
ggttaggagt ttgagaccag cctgaccaac atagagaaac cccgtctcta ctaaaaatgc 12540
aaaaatagcc gggtgtggtg gcacatgcct gtaatcccag ctacttggga ggctgaggca 12600
ggagaatctc ttgaacccag gaggtggagg ttgtgttgag gctaggttgc gctgttgcac 12660
tctagcctgg gcaacaagag tgaaactcca tctcaacaac aacaacaaaa aagaagacga 12720
caggctgtgt gcagtggctc atacctgtac accagcactt tggaaggcca agtgggaagg 12780
attgcttgag cccaggagtt cgaaaccagc ctgggcaaaa tagcgagatc tcgtctctac 12840
aaaaaaataa aaaattagtc aggtgtgttg gcacatgcct gtagtcccag ctataggctg 12900
aggcagaagg aatgcttgag ctcaggagtt ggaaggtata atgagctgtg atggcaccat 12960
tgcactccag cctgggcaac agagtgaatg ggggcatggg gaaggccagg tgtggtggct 13020
cacacctgta atctcagcgc tttggtaggc tgaggcggga ggattgcttg catccaggag 13080
tttgagacca gtctaggcaa cacagtgaga ctcccatttc tacaagaaaa tatttaggct 13140
gtaagtggag agttattgtt tagtgggtag agttttctgt tttataagat gaaaagttat 13200
ggagatggat agtttaatgg ctacataact ttatgtaata ccactgaact atacactgaa 13260
aaattggagc tgggcttgtg cctgtggtcc cagttaatca ggaagctgag gcaggagggt 13320
tacctgagcc caggagttcg aggctgcagt gagttataat tgagccactg cattccagcc 13380
taggcagcag agtgagacct tatctcttaa aaaaaaaaaa aaaaaggaaa aaataatagg 13440
ctgtaaacag ttgattcctt accagtaact tttaatattg aattaacaaa tatttatggt 13500
atttttaaca tgctcagagt tcaggctttt ttcagctgtg atctaaatga aagtgagagt 13560
gaggtggtga gagaagtcag tggaaatgtg agtgaaagat tgaggaagag ctggcttaga 13620
tagccacagt gtgtagttat tttttttttt agagacgggg tttcaccttg ttgcctaggc 13680
tggagtgcag tggctattca caggtatgat catagcacac tacagccgca aactcttgag 13740
ctcaagccaa cctcctggtt cagcctccca agtagctggg accacaggct catgccacca 13800
```

FIG. 3E

```
catccagtca aagtgtctgg gttttttggt tttcaattag gttcctcagg ttaaatggaa 13860
gtatgtggtt ttttgttttt ttgttttttt tttttgtaaa gacagtgtct cattatgttg 13920
gccaggctgg tcttgaactc ctggccccaa gccatcctcc tgcctcaccc tcctaaagtg 13980
aagtgctgga ataacaggca tgagctgcta ggcctggcca aagttgaaaa gtaaagtgat 14040
aagggagaag taaatgtatt ttatcctctt ggagaatttg cttgacctat tgatttttat 14100
attcaaattg gccttttgaa ctcttagatg agttccttat tttgacaatg aatgcgtgtg 14160
tccctatctt ttaattaaat tgccagccta gagtgtactc ccaggggcca aatgagtaat 14220
tatcttgtta cttaatctag aaagcatact gtaaggtaac taaaggatct gaaagtgtgg 14280
gaatgaagaa agcaaccctc caaatatatt tgcttctaaa actcttaaaa ttttctgaac 14340
actcataaag ttgaatggct ttaagaagag tgagacttta taattgaccc agttagttaa 14400
agttacatgt aattttgtat aacagtttaa accttatttt tgtggaagaa aaagaagcag 14460
tttaccataa ttgagtagcg aggccttaag ggtaagaaat accatgcttg atacctagtt 14520
tcttcttttt ttttttttt ttttttgag atggagcttc gctcttgttg ctcaggttgg 14580
agtgcagtgt cacgagaaaa tacatttaca gtgctatact gtattttctt tttctttttt 14640
tttttttgg ctttgagatg gagtcttgct ctgtcaccct ggctggagtg cagtggcacg 14700
atcttggctc attgcaacct ctgcctcctg ggttgaagcg attctcctgc ctcagcctcc 14760
cgggtagctg ggattatagg cggcgccacc atgcctggct aattttgtat ttttagtaga 14820
ggcagggttt ctccatgtca ggctggtgtc gaactcccga cctcaggtga tccgctcgcc 14880
tcggcctcac aaagtgctgg gattactggc gtcagccacg gcgcccggcc tctagtttct 14940
tctatcagtt agcaagtatg ttttaagctc aacggggaaa ggagaaatga tgagtctctt 15000
acccctagga gcctacagtt aagggataag acatgaagca attatgaggt cagcatgaaa 15060
tagtacataa tcaattatta aatggttttg tatatagttt tgaagttgtg tgatcagttt 15120
gtgatagggc agtgaatcac aaatgtactt tttagcttct gtatcctttg gataatctta 15180
cttggaactt gagtataaaa aatgaataaa aacaacagag ctgccctgat taaattgagt 15240
ggtggagctt ggagtcaaaa agatagcctt ctgtattcac aggttccaca tcctctgatc 15300
catccaatca cagaaaatat ttgggaaaga aaacaataaa aaataataca agagctaata 15360
caaattaaaa aaatatatag tgtaactatt acataacatt gtattaggta ttataagtaa 15420
tctagcgatg atttaaagta tatgggaggg tgtacaggtt atgtgcaaat atgccatttt 15480
atataaggga cttgggcatc tgaggatttt ggtgttcaag gattttggag tcagtcgggt 15540
cggactaatt gttagatggt ggacatgaat ttagaggtca ttgagttgtg taattttgta 15600
cagtaaaagt atgtgtgtgt aaattttaat taaaatgtat tatacagaaa ctagtgtatc 15660
tttaacagaa tatcttgagt ttcttttgat tgactgattg attggtcgat tgattgagat 15720
agggctctcg ctgtgttacc caggatagca ttgggctcaa gtgatcttcc tgcctcagcc 15780
tcccaaagct ctgggattac aggtgtgagc cactgcaccc agcaggtttc ttaattacgt 15840
aggctagtgc acttgcaaaa tatagccaca actataatat agtctgtttc caagccccca 15900
aaaggctccc tcctgccccc tccaacagtt agtacctttt ttccaaaggt aactgttggt 15960
ccttatgacc tcctatcacc catatttgtt ttgtgtggtc atgaatttca aactgaataa 16020
cattctagca ttctttgtga aatttgtgtt cttacgtaat ggcagtagtt cacattcttt 16080
ttcattgcta tgcagtattc aaaattgtgt aaatatacca cattttactt attcattcca 16140
ttatgtgtat ttgggtagta tcaatttttg gctgtataaa taaagatgct gtgaacatta 16200
tacctgagca cctattttga gtatatatcc agggatagag ttactggata atggagccac 16260
atgtatattt agtttttcct gcagcatttt actacagaaa ttttcaaaaa taaacaagag 16320
ttgacagaat tgtttgaata cccatattca ttcctaccac ctagattctg cagtaaacgt 16380
tttgctgtat ttgctttcac acatacctat tcgtcgttct tctgtcaatc tatcctgttg 16440
atgtgtttca aaaataagtt gccaaatagt tttccaagtg gttgcactaa ttttctgctt 16500
gtactgacaa tatgagagtt tcatttgtgc cacgtcttca ccagtatctg gtattgtcaa 16560
tctttatttt tgaaactttt tttagctttg aagtataatt gatataatta tatatactta 16620
tgatgtacac catggtgttt taatatatgt atacatcatg aaatgattac cacagacaag 16680
ctaatttaca gatccttcac ctcacagtta tgatttgcta ttgtgagaat atttaagatg 16740
tattttcata gcagttttca agtatacaca atacattatg ctaagctgta gtcactatgt 16800
tatacaacag atctctggaa ttgtagatct gttttatttt agacatttta ataggtgtgt 16860
agaagtatag aaatgcagtt gacctttgaa caacatagag ttgaatagcg tgggtccact 16920
tacagtgcta atttgtttac acaagtatgt tggaattttt gtttgagatt tgtgacaatt 16980
ttttttttt acgcttaatt cgctttattt tccttgtcta aaaaccctat gttgtagcca 17040
cagctggagc ctgggtccgc tgcacggaga ctctggcgtg ggccttgatg aggtggtcag 17100
tgaattcctg atcgggagac ttggtaaata cagtctcctt ccagaggtcg ggggtcaggt 17160
agctgtaggt cttagaaatg gcatcaaagg tggccttggc gaagttgccc agggtggcag 17220
tgcagcccca ggctgaggtg cagcagtcat cgataccagc catcatgagc agcttcttgg 17280
```

FIG. 3F

gcacaggtgc ggagacaatg ccagtgcccc tgggtacagg gatgaggtgc accagtgcag 17340
agccgcagcg gcctgtcacc ttgcaaggga cggtgtgggg cttgccaatc ttgttccccc 17400
agtagcctct gcgcacgggg acaatggaga gcttggccag gatgatggcc ccatggatgg 17460
cggtggccac ctccttggag cacttaacac ccagaccgac gtggccattg tagtccccga 17520
tagcaacaaa cgccttgaac ctggtgcgct ggccagcacg ggtctgcgtc tgcactggca 17580
taatcttcaa aacctcatcc ttgagagagg cccccaggaa aaagtcaatg atctctgatt 17640
ccttaatggg cagggagaag agatagatct cctccaggga cttgatcttc atgtccttca 17700
ccaagcagcc cagcttggtg acgggcatcc actccttatc ctcggccttg cctccgcgag 17760
ctccgcggcc tcggccccgg ccccgtctac ggctgcgacc ccggccccgg atgccactgc 17820
cgaaacctcc gcggaagcca ccgcagttcc ccatcccagg gcctccgggc ccccccccgc 17880
tgcactggtg tcatccacca tttggtgttt tctcggagaa gaaaggattt gtgacaattt 17940
gaataaactt aggtgattct catttttga aatatcaaaa aattaaggta tgtcatgatg 18000
gatgcataaa atatgtaaat actagtccat tttatcattt actaccgtaa agcattcaca 18060
tctattataa aaaattaaaa tttgggctgg gcatggtggc taatacctgt catcccacac 18120
ctttgggagg ccgaggcagg aggatcactt gaactcagga ttttgagaca gcctgggcag 18180
gagagcaaga cctcatcttt acaaaaaatc aaaagaatta gctggagccg ggcgaggtgg 18240
ctcacgcctg taatctcagc actttgggag gccaaggtgg gtggatcact tgaggtcagg 18300
agtttgcgac gagcctggcc aacatggtga aaccctgttt ctactaaaaa tataaaaatt 18360
agctgggtgt ggtagtgcat gcctgtaatc tcagctactc aggaggctga ggcaagagaa 18420
ttgcttgaat ccgggagaca ggtcgcagtg agccaagatc atgccactgc actccagcca 18480
gggtgacaga gcaagactcc atctcaaagc aaaaaaaaaa aaaaaaaaaa agaaacagaa 18540
aaagaaaata cagtataaca ctgtaaatgt attttctctt ccttatgatt ttcctaatat 18600
tcttttctct ggcttatttc gtaagaatgc aatatgttat acatgtaaca tacaaaatat 18660
gggttaatca agtttgtgtt atccataagg tttctttggt caacagtagc ttattagtag 18720
ttaagttttt gggaagtcaa agtttaggcc atgcatgggg catatacccg taatcctagg 18780
tatatgtctt gggaagccaa gacagaagcc aagacagcag atcatttgag ctcaggaatt 18840
cgagaccagc ctgggcaaca cagcaaaact ctgtttctat aaaaaataca aaaattagct 18900
ggatgtggtg gcgggcacct atagtcccag ctacctgaga ggctgaagca ggaaaattgc 18960
ttgagcctgg gaggcggagg ttgcagtatg ccaagattgc gtcattgcac tccagcctgg 19020
gtaacgggag tgaaaccctg tctcaaaaaa aagaaaaatt atatgaaggt ttgcgactgc 19080
actgggggtt gagacccctа accccatgt tcaagggtca actgtatttc attgtggttt 19140
taatttgcat tttcccagtg agtagtagag ttaaccttтt caagtgcttt ttggcctttc 19200
gggtattgtc ttttgtgaag tcttttgtcc agtttgtgtg tgtgtgtgtg tgtgtgtgtg 19260
tgtgtttggg ggggacagag agtcttgctc tgttgcccag gctggagtgc agtggcgcga 19320
tcttggctca ctgcaacctc cacctcccag gttcaagcga ttctcctgcc tcagcctccc 19380
aagtagctgg gattataggc gcccaccacg cgcggctaat ttttttgtag ttttagtaga 19440
gatggggttt caccatgtca gccaggcttg tctcaaaact cctgacctcg tgatccgccc 19500
gcctctgcct cccaacgtgc tgggtttaca gatgcgatcc accacgcccg acgttgtcca 19560
gtttttaatt gggttgtctg tctgttaaaa acttgatttg taggagttta aagtatattt 19620
tactaataaa aaagaatggt aattgttctc tgcataatat ttagaaaata aggccgggca 19680
caggggttca agcctgtgat ttcagtgctt tgggatgcca aggtgagagg atcccttaag 19740
ggaaggaggt tgagaccagc ctgggcagaa tactgacacc ccatctctac aaataatata 19800
aaaaagttag gtgtggtggc atgcacctgt agtcccagct acttggaaga ttgaggcagg 19860
aggatcactt gagcccagga gctcgaggta acagtgagat atgatcacac cactacactc 19920
cagcctgggc aacctcaaga tcctgtctct taaaaaaaaa aaaaaaaagg aggtggggga 19980
aacgaaataa agaatagaag gaaagtagag attgtgttag atcctagaat gtgaagataa 20040
gtaaaatgat ctccatgttc tctgtacccg cgttctagag agggatgtag gtaaatggat 20100
ataaaataca ctgccataaa ttttatgaag aagaatgttt agggtgctat aatttgctca 20160
acacttcatt atttttgctt tcttgatatt ttaaataata ctatgaatgt ttaatcgata 20220
aaacacttct tgttttagaa tattttattt gtataaattt ttttaggaat tgctggatct 20280
tgccaggcac aatgactcat gcctgtaatc ctagcacttt gggatgtgga ggcgggagga 20340
ttgcttgagc ctagaaattg gagactagcc tggacaacat agtgaggtcc catctctaaa 20400
aaaaaaaaag ggaaggaatt gctaaatcaa aggatgtgag gattttaaag ctctacacga 20460
tcaatcggat catcctaaat aatttatcga aacagtggaa aaatatgttc cacaataaag 20520
catacctctt tggtgaccaa aatgtggcat tgatgaatat tgggggttc cccctaaaac 20580
tttaaccccc tttgtgttgg cacattcatg agataattct gttgattagg ttatagtttt 20640
ttttttaaat caaagaataa aacatggtct ccttgattag agttagatat tccttgcttc 20700
cttgaatgaa ttccttttttc tttttctgtt tttgagatgg agtcttgctg tgttgcccag 20760

FIG. 3G

gcctgtagta cattggcacg atctccactt actgcaacct ctgcctccca ggttcaagca 20820
gttgtgcctc agcctcccga gtagctggga ctacaggtgt atgccaccat gcttggctaa 20880
ttttttgta tttttaattt ttttttgaga tggagtctcg ctctgtcacc caggttggag 20940
tgcagtggca ggatctcggc tcactgcaag ctccatcccc tgggttcaaa cgattctcct 21000
gtctcagccg cctcccaagt agctaggatt acaggcgtgt gccaccactc cctgctaatt 21060
tttgtatttt tagtagagag gggatttcac tatgttggcc aggcaggtct cgaactccca 21120
gactcaagtg atctgcccac ctcgtcctcc caaaatgctg ggattacagg tataagccac 21180
tgcacccggc catttttttg tatttttgta aaggcagagt ttcgccacat tggccaggct 21240
ggtctggact tcaggttgcc ctggcccctt aaagtgctgg gattacaggc atgagcaccc 21300
agcctagttc ctttttcaaa tgaatagcac ataagtatat gctggctctt actaacttta 21360
ctctgggaga aaaaaacata gctccaattt tcttttcatt ataggctctt tcttgaatag 21420
aatacttgca tagaaaagca atagcttata cattggtgag gggagtgttt acattgcccc 21480
caacatactg tttaatgcat aacatacatt attttattta attatcacat agtatctgta 21540
aagaggaact atattatttt taatactatt aatatgctat ataatatgta tatatatgtt 21600
tctactactt atatatcatg caatgtgtac attatagctg acatagagac tcagcagttt 21660
gtactctgga taagtttgtt acatttaatt ttttatttt ctttttagtg acagggtctc 21720
acactgttgc ccatgcttga gtgcagtggc acaatcatag ctcattgcag cctcaaactc 21780
tgggctcaag cgatccacct gcctcagcct cctgagtagc tgggactaga ggtgcatgcc 21840
accacacctg gctaatttttt gtctttttgg cagagacaag gtctcactct ttttcccagg 21900
ctggctttga attcctgggc ttaagcaatt ctaccttggc ttcccaaagt gttgggatta 21960
caggcgtggg cgactgcacc cagctgtgtt tcactttctt gctggtgttt ttggaagcac 22020
aaaagtgatt aatcctgata aaaaccagtt tgtctatttt tttctttggg gagcttttgg 22080
tatgttcaag gaggggggtca aacttgattc ttctgcatgt gactagcctg atgttccagc 22140
accatttgtt gaaaatagtt ttctttttctc attgaattat cttggtattc ttgccaaaga 22200
tcagttgtct ataaatatga gagtacatct ctggactctt ctgttccatt ggtctatatg 22260
tctgtcctta ggctagtact gtactgtagc tttataagta aggtttgaaa tcagaaagtg 22320
tgagtctttt ttgttgttgt tgttctccct tttcaggatt gttatggcta ttcttagtcc 22380
cttgtatttc catgtgaatt ttaggatcaa ctgtcaattt ctgcaaatga gccagctaga 22440
attttgatag agctacattt aatttttttt aatgaaccca cattgaaata ttcaacgatt 22500
gccaaaacat tatcacccca aacccactgt ttatcttaga gttcattctt ggtgttgtac 22560
atggatttgc catgtataat gacatgcatc caccattata gtatcatacg tgtagtttta 22620
ctgctctaaa aatttccgt gttccatctg ttcatccctc cccccaacac ctgacaacca 22680
ctgatggttc tattgtctcc atagttttgc catttgcaga atgttgcata tttggaatca 22740
tacagtatgt agccttttca gattggcttc tttcatgtag taatatgtat ttaaagttcc 22800
tccatgtttt tcatggcttg atagctcatt tatttttttag tgctgaataa cattccattt 22860
gcattgatat actgtagttt atctatttac ctatcgaagg acaccatagt tacttccaca 22920
tttcgatagt catgaataaa gctcctgtaa acatttgttt ctgggttttg gttggatgta 22980
tttttcaact cctttgggta aataccaaag ggtgcaattg ctggatcaaa tggtaataat 23040
atgtttagtt gtgtgagaat tatcaaactg tcaccctaag catacctttt ttgttgttgt 23100
tgttgttgtt attgtttttt gagatggagt ctcgctgtca cccaggctgg agtacagtat 23160
ggcgtgatct cagctcacca caacctccac ctccctggct caagcgattc tcgtgcctta 23220
gcctcccgag tagcttggat tacaggcacg tgccaccatg cccagctaat ttttgcattt 23280
ttagtagaga cggggtttca ccatgttggc caggctggtt tcaaactcct gacctcaggt 23340
gatcctcccg ccttggcttt ccaaagtgct gggattactg gcgtgagcca ccgtgcctgg 23400
tcgcaaagca taccattttg cattcccacc agcagtaaat gagagtttct gttaccacac 23460
atccttgcca ggacttaatg tcagtgtttt ggattttcac cattctgata ggtgtgtgtg 23520
gtatttttg ttttaatttg cagttctcag tgcttctcag tacctcccac ccaccttatc 23580
tgggacagga tggctggagg ggctggagtt gggtagttct ctttccccag gtaggttaag 23640
ctctggtaaa agtttctccc agaactttat attaaattaa aagtttcgca tgaggataag 23700
acttattagg agcagaatgc tctggcatat ttaaaatggt ttattttcct ctccgcctcc 23760
tggaagcact cacgtgagga cctggtagag gttttggagg taaaattcac aaaagtgtgg 23820
tggccccaat tgggtccact tggagttttt attttaattt tctttatttt ttatttacat 23880
tttttgtttg tttttttgtt ttggagatag tcttgctctg cggaccagaa tggaatacaa 23940
tggcacaatc tcagctcact gcaacctctg ccccctccga ttcaagctat tcttgtgcct 24000
cagcctctct cgagcagctg ggattacaga tgtacaccac catgcccgac taatttttgt 24060
atttttttgta gagatggggt ttcgccacat tgcccaggct ggtctcaaga tcctagcctc 24120
aagtgatccg cctgcctcag cctcccaaag tgctgggatt acaggcgtga gctactgtgc 24180
ccagccttga ctttttaata atagccattc tgactgatga gatggtatct cattgtggtt 24240

FIG. 3H

```
tttgatttgc atttctcttc ttatttcttt tcttttcttt tctttttttt ttttttttt 24300
gagacaaagt ctgttgctga tgctcaggct ggagtacagt ggtgcaatct tggctcactg 24360
aaacctccgc ctcccagggt caagcagttc tcctgcctca gcctcccgag tagctgggat 24420
tataggcacc caccaccatg cccaactaat ttttgcattt gtagtagaga tggagtttca 24480
ccatgttgcc caggctggcc tcgaactcct gacctcaggt gatccgcctg ccttggcctc 24540
ctaaagtgct gggattacgg gtgtgagcca ctgcacccag ccaatttgca tttctctagt 24600
gatcagtgat aatcggaatg gcacgcaatt tacaacatta gttgtttatt tctggaattt 24660
tccatatagt attttcagac catggttgac cacagttaac tgaaaccttg caataagtgg 24720
ggactgctgg agatcctctt cattttaaag atgcagaaat aggttgggca gtagagcagg 24780
gcaactagga gaaatgcatt taaacaatag ggataactgc acagtgaggc agaaaatgtt 24840
atataataca ctagatttgg ctcaaaggca gtcacaggag cctctaggag ttcatttcca 24900
gtagagtcca ggagaggaat gagatactgg aattaagtag gttatacagt gatacttgaa 24960
gtaatgagac ttctttaaaa tgaaatgggt cttctaaaaa tgtttcataa atcctaaagg 25020
taggcacaac atatggcaca taaattatca gaaataccca tatgtataac atctcttttt 25080
tcatttggaa gtaaacagcc ctgcctcagc aaatttctcg agttaataaa atgtttttag 25140
tgtccccgtg tcttcttttc tattctcagt tggtgatttc ctttcttttg ccaatctccc 25200
tgataggcaa tctgtgctct ctgtctaacg tttctctctt gaatttatag tgtagttatt 25260
actaaggcac caattcctat caattaaggc tcatttcagt gtaatgtgga tgtgttttat 25320
ttagtattat gtccactggg cctttgtata tttagcgttt gcaccactcc ccaccccatt 25380
ttcatgtgac aactggaggc cttgttcaaa aaactaccag aaaaaggaca tgtatcccat 25440
cctttatggc tttcttgtaa caccagaaac tatccagcct tgttttgtcc tgtttactgt 25500
aagatcctaa gccagtggtt ctcaaccagg atgaatttgc tctccaggaa acgtttagca 25560
aagtctagag acatttttttg gttgtcacaa tgtggcaggg aagagggtgc cactggcatc 25620
tagtggatgg agagaactgc caaacaccct gtgatgtaca ggacagcccc tacaacaaag 25680
aattgtctta ctaaaatgtc agtattgccc aggttgagaa gcccttcact aagcaagccc 25740
cttttcctct ctaacctcag ttaggtcctt aattaagttt gtttcttcaa ctgaaaagtg 25800
gaggttattg caacttttttc aaatggttgc aaagataaat gaatttaagg cgtggaaata 25860
tttagcacag tatctgacac acagtggcac tctgctgaaa gaataaatta gggtttttaa 25920
aaaatgttgt ataatgcagc ttctcatcca tcctgatttt cctgcaaaaa ttctcatgtt 25980
aaatgttttg atccattatc actgtaaatc tcttacatgt tatagaaatt cctgtattta 26040
agcttctaaa atttcctaga ctacctctta aatatgtttt gatttttggt catgaaagca 26100
atgattatac cagcactgga acaaatcttt ttaagtaaca gaaacctttt cctgttcttt 26160
tctaaagcac ttcagtgaat taaagattgg tgaggagtgt aggtcaaagt ttaagtcaat 26220
cagtaattga atattttctg ttgtcaaaga gttaataagt tcatgatcaa atagcgtatg 26280
taggagaaga cgttagaaag atttcagaaa cactaaactt atcgcattct ctttgtgctt 26340
ttaattttat gtagcaattc aaggccatca atagttgtag tcttaaaagc ccctaaaagc 26400
agaaatgata ttactttgtg gattgtttag atttttatga tgtacccctg tttgcaaggt 26460
ggtgttaggt cctatgaatt aggagtttgt aagtttagtt aattggctga ggcccattgt 26520
tgatctgtta atttttttgt gaattagttt ttaacatacg gatcctgcac aagataaatt 26580
tatacctttа aacatttcag ttttttggta tcgtaaatgg tgttcaaact ttttattttg 26640
gcttccagtt cattgctggt atttacaaat acaattaatt tatgtgtgtt cttttaacct 26700
gcaaccttgc tgaattcatt ttaggaggtt ttggttagag tccttgggat tttctaaata 26760
ggctcatgtc tgtttgtctc tctttaatta aaatttttta gggtgtgttt ttgtaaagcg 26820
ttaatattca gttttgggtg ttttcattct cagactgatt ctcaacctgg ccctctttca 26880
aactttcaca ccttctccct tttcaacttc actcccttcc ccagaaatca ctgaaatgac 26940
ttctgagcac attgtcattt tatcttactg attcaaaatt attattagat acgttattat 27000
accctggagg taacagaata gtgcatccag agtctttaat tatgtaggac aattgtgtta 27060
gtgggtgaaa gctgttaggc ttttggattc ttgcacctac tttttattga ttttttttтc 27120
ctcattgtca gttcttcatc tgctgttagt gtgtctgccc ttaaactacc tcttcctact 27180
tctaataatt tggcatgttg gaaaaccaga tactcaagct catcacagta ttttctagca 27240
actcccaata gagagatccc agaggcagtg gcagttaagt tgcattttca cttgtgaata 27300
aactattctg gggtttagat acataggtac tttatatttt tttaagttaa aaatttttta 27360
aagtttggaa caaatttgct taataaagtc agttttagcg ggaaattcac atgcacatga 27420
gaatcgaaag aaagaacctt taatgttgat gtaagcttat cattcctttt tcacatatac 27480
ctcaccacct cctctgtgtt taattatctg gacttttttc tcatccctag attttgtttc 27540
tcaaaaatgt ttctcccttа aaacatgttt taatgaactc tgtaagatct aagttcttaa 27600
tctcaaagta atctttagat tgaggagtac taaaattgag atccacagac cctttaaggt 27660
cagggataga actgctttca gtgtgattga tttccttagt aatcttatgt gttttattca 27720
```

FIG. 3I

aaaacgttat ctgaaatagc tttcattgga ctgcataagg tttccatggc aaaaaaaaaa 27780
agtgttaaga atccgtgaga ggcaatccca gagtgagctc ctgtcattgc tgccagaaac 27840
cacaaagcct tttgagagtt ctgaactctg agtcacagag ttctgatagg tatgtggtat 27900
tgtaattcag ttgaattttg ctgttggaat atttaggtaa attgtcagac gtaattaaga 27960
tttgcagcat gaaatgggcc caaatatata tggagtcttg ctctgttgcc caggatggag 28020
tgcagtggca cgatctcagc tcactgcaac ctccacctcc caggttcaag tgattctcct 28080
gcctcagcct cccaccacgc cctgtctgtg ataagatttt ttttttttct tcttgagacg 28140
gagtcttgct ctgttgccca ggctggagtg cagtggcacg atctcagctt actgcaagct 28200
cgtcctcccg ggttcacacc attctcctgc ctcagcctcc ggagtagctg ggactacagg 28260
cacctgccac cacgctcagc ctgtgataag atttttaaat aaaaatttca cggctcgttg 28320
ctatggctca tgcctgtaat cccagcagtc tcggctcact gcaacctccg tctctccggt 28380
tcaagcgatt ctcctgcctc agactcccga gcagctggga ttacaggcac ataccaccac 28440
ccccagctaa tttttttctgt tttagtagag atggagtttc actatgttgg ccaggctggt 28500
ctcgaactcc tggtctcagg tgatccaccc gccttggcct ccctagtagc ctgaggctac 28560
aggcatgtac caccactcct ggctaatttt tgtagggatg gggtttcgcc atgttgccca 28620
ggctgggatt gcgggtgtga gccactgcac ttggccccaa aacaattttt aacggatgat 28680
gctgggacac ctacatataa ccacatagtt aagtgaagtt gggccccttt cttactatat 28740
gcaaaaatta actcagaatg gttcatacac ctaaatgtaa gagctaaaac tttataaacc 28800
tcttagaaga tataggagca aatgctcatg accttgggtt tgacaaactt cttaggatac 28860
tgaaagtaaa agtaaccaaa gttgaatttc attaaaattt aaaatttgtt cttcaaataa 28920
cgtaaagaaa gtgaaaagat aacccatagg atgcaagaaa atatttgggc tttttttttgt 28980
ttgtttgttt gtttgagaca gagtctccct ctgtcatcta gactggagtg cagcagcaca 29040
atcttggctt attgcaatgc cacctcccag gttcaagcag ttcttgtgcg tcagcctccc 29100
aagtagttgg gattacaggt gtgcactacc accgcctgct aatctttttt ttttttttt 29160
tttttttgt atttttagta gagacaaggt ttgaccatgt tggccaggct ggtcttgaac 29220
tcctgacctc aggtgatctg cctgcctcag cctctcgaag aaaatacttg taaatcattg 29280
gtctgctaag ggacttgtgt ccaggagata acaaaaagat aatttgctat cagagagggc 29340
aaatcaaaat cacaatgaga tatcacatct tatccacttg aatggttata ataaaaagat 29400
taagtttaag caatgctgta gagaaactgg aacattctat ttccctttt aaaatgcaca 29460
cttcattgat ttccagtgtg ttcagagtta cacaaccatc acttaccaat acagctctcg 29520
atgggattgt aagatggtgc aggtgctttg gaaaagtttg gcagttctac aaaaggtttt 29580
gcagagttcc attcattggt ttatataccc aataaaatga aaacatgtac aaaaacttat 29640
acatgtattt tcataacact gttattcaca gtagctaaca agtagaaaca acccatgtcc 29700
atcagtgact aaacaaaatg gtggtggact attcagtgat aaaaaggaat gcaattctga 29760
taaatgctgc atcatgggtg aactttgaaa acatgctaag tgaaagtcga gccacaaaag 29820
accacatatt agtatgattc catacacagg cataccttgt tttattgcac ttcacagggt 29880
tgtgtgtgtg tgtgtgttta cacaacttga aggtttgtgg cagctctgca tgaaacaatt 29940
cttttggcac cattttccca gcagctcatt tcatttctct gctacatttg gtaattctcg 30000
tgctttcaga ctctttcatt atatctttta cagtgatctg tgatcagttt tattagggtg 30060
gtctggaacg aaacccacac atctgaataa tgcttatcta tgaaatgccc agaataggca 30120
aatctataga gacaaaaagc aaattagtta gtagtagcct agggccgggg agtttggggg 30180
agtgactgtt aatgagtaca ctttcttttg ggggtaatga aaatgttcta aaattgataa 30240
gtgatggttg catagctctg aacacactgg aaaccattga agtgtgtatt ttaaataggg 30300
aaatagaatg tgaattatat ctctaaggtg ttacagttaa tggaagcctt tcaccttatc 30360
gtgatactgt atttcctcag tgctgagatg tgctttttaa ggctaggctt ggctgctcac 30420
tcctgggatc ccagcacttt ggaaggccaa ggcaggagga ttgcttgagc ccaggaattc 30480
tagaccagcc tgggcaacat agtgaaaacc tgtctctaca aaatgtatat ttttaattag 30540
cggggcatag tagcatgggc ctgtagtccc agctactcag gagactgaag tgggaggatc 30600
acttgagccc ggggggtcgt caaggctgca gtgagccaag attgtgccac tgcactctag 30660
cctaggtaac agaatgagac cctatctcaa aaaaaaaaaa aaaaaaaaaa aaaaaaatat 30720
atatatatat atatatatat atatatacac acacacacac acacacacat ttatcaggaa 30780
gatagatgtg atgctagctt ttacatctct ctgtgtaatt gttcaggaac tgcatcagtg 30840
ctaatctggg ggtttttgtt ttggtcaaga tttcatggaa catcagaagt agaacataca 30900
ttgagaccaa ctattctaaa tcctccattt tatggtctta aatgtcagtc tggcaaaatt 30960
cagagaaaac atttcccagt aggagactgc agctgggaag aaggtggtgc cagcagcagc 31020
ttgtaaggat ataaagagaa ggatcagaag agctggctct tatttcctac tggaggaaaa 31080
gataccgctt ttggtgacaa gctgcatgct ttagcagtga ctttatttttt ccttcaccac 31140

FIG. 3J

aaaccagact tgtgctattt ccagcctcag cacattattt aggcacttcc ccactgcatt 31200
gttttcgttt ctgaaattgt atcacattcc ctaaattgaa tccaactttt aatgtagtgc 31260
ccttcagaca gacacgtttt acattcacct gttttgtgtc tttcttggct taaaccacct 31320
aagattaggg cccgatacat tcttgacaac atccatcatt ttataggtgc ttcatttata 31380
tattaacaaa caaccggaat gcttcagaga tttttcgcat cagatcagac cttgctgttt 31440
ttgatgtttg accctatctt ctggccgggt gcagtggctc acacctgtat cacacctgta 31500
atcccaacac tttgggaggc caaggcaggt ggatcacttg aggccaggat tcgagaccag 31560
cctggccaac atggtgaaac cctgtctcta ctaaaaataa aaaaattagc tgggcgtggt 31620
ggcgcgtgcc tgtactccca gctactcgag gttgaggtgg gagaattgct tgaacctggg 31680
aggtggaggt tgtagtgagc caagatcgcg ccactgcatt ccagcctggg tgacagagtg 31740
agaccctgtc tccaaaaaaa aaaaaaaaaa gactgtatct tctgtcctgc ttcagcatat 31800
tttataaccc tgtcttccat ttctattcta ccccccttcca tgctgctgca tttttccaca 31860
atgcatattt taatttttaaa ttattccttt tctccttttc ttaaaggcag atgatttcct 31920
ttggattctc gctctttctc tcttttgcat gcagacatat ggagaactaa agaaagtatt 31980
aaggaaaaga ttgtttaaat ttatttgaaa tttgcttctc tgtgaattcc atcttgaaat 32040
ttaaactcac agggcaagcc catgtgtctc ctaatcaaca acgaggaaat ggtagttagg 32100
taagactgta taagacattt aagtagatga tagtggaaaa gggaagagag gaagtcagag 32160
gaatcaggaa ggatgagatg tctatctcca gtgactagtg gtgtcattag agaagagccc 32220
atatattaag taagccctta tcttggtaag tgctttactt atctgcttat tccttacaat 32280
actttttttt ttttttttttt aacagagtct agctctgtgg gccaggctgg agtgcagtgg 32340
cgtgatctca gctcactgca acctccgcct cccaggttca agcaattctc atgtcgcagc 32400
ctccagagta gctgggataa cagagctaaa tgactgcatc tggcctccaa ccctttatta 32460
agataggtat tttgtctgtt atgctgatga ggaaccttaa tgctcagaga tattaagaag 32520
cttgcgtaag gccacataca taattaaata ccagagcaga aattccaggg aatgcctgta 32580
ctctttccat tatgtatttt tgttcattaa cctaggtagg gggcatagaa attctgagtt 32640
tgagaggcat gtggaaggaa gtatatttgg tttggttttg gacctattga agttcctgtg 32700
ggatgggtag gtggagtttt gaaagagaca acaacccagg tattttcact tgggagaatg 32760
ttgggtgtaa tttatatcaa gttttccttt ttggttttat tgttatttaa tccaaatcat 32820
tagcagtttt tctttttttt ttcttttga gacagcagtt tttcttttc ttattgttat 32880
ttaatccaaa tcattagcag ttttttcttt ttcttttttct ttcactttgt cacccagtct 32940
ggagtgtagt ggtgtgatct cgtctcactg tagctcaacc tcctgggctc aagcgatcct 33000
gctgcctcag tccccgaggt agctgggact acaggtgtac atcaccactt ctggctaatt 33060
tttgtatctt ttatagagac ggggtttcac catgttgccc aggctggtct caaactcctg 33120
agctcaagct atccgcctgt cttggccttc caaactgcaa ggattacagg cgtgagccac 33180
tgtgccaggc gcattagcaa tttttcatag catatctaca aagttctaat atgtatataa 33240
tgtatattac aagaacataa gtgctttgat ttaggttatt cagttatata attatgttga 33300
gtcccaactg tgtgccagat accgtgctga gtaggtagta aacattggta agcaaaacag 33360
acatagcccc tgcccttatc agactaccca catagcccgg acacagtggc tctcacctgt 33420
aatctcagca ctttgggagg ccaaggctag aggatcgctt gaggccagga gtttggtcag 33480
cctggtcagc gagacactgt ttctaccaaa acaaaaaaat tgctgggcat ggtgttgtgt 33540
acctgtaatc ccagctactg gggaggctgc agtgggagga ttacttgagc ctagaagttg 33600
gaggttgcag tgagatgtgt ttgtatacct tttgtcatta ggtccatatt taggccctga 33660
aataggctaa tcaaaatatt tatttatttt tcatttgttt tcagctgggg acgaaacaca 33720
gtagctttgc aaatataaag tgggagcaca acctagcact ggcgttggag gtcctcaaaa 33780
cagtccccag gtttggtgat tcaataggag ggctcttagg acttagtata taatcttact 33840
cacatctatg atttgttaca gcaaagaatc aaagcaaaat cagctaaggg gaaaggtgct 33900
ttggtctaaa gggaactaag cacaagcttc caagagttat ctcctagtgg agtcacacag 33960
gacatgctta attccttcaa catatatgaa aggttttctg ccagagaagc tcatttggag 34020
actcagtgcc ccaggtactg attaggcacc cactgcctag catatgacaa aattcctgat 34080
tctcagaggg gaagcaggtg ttgaacataa accatatagt gttcgcagat agggaatggt 34140
gggaaccacc cccaaatcca tatctttttt tgttgttaag atggagtctc tctctgttgc 34200
ccaggctaga gtacagtggc aagatcttgg ctcactgcaa cctctgcctc ccgggttaaa 34260
gcagttctct tgcctcagct tactgagtag ctgggattac aggcgtgcac caccacgccc 34320
agctaatttt agtattttta gtagagatgg tgttttacca tgttggccag gctggtctcg 34380
aactcctgac ctcaggtgat acgcctgcct cggcctttca aagttctggg attacacgca 34440
tgagctacca tgcccagcca tcacatcttt taaaattgtt gtttttatac ccctgtctaa 34500
tcttctactc actcagtgtc tgttagccaa caattgaatc aggattccat tttgctgcct 34560

FIG. 3K

```
agaaaatcag tggctgtcta gttatgaaac aattgataac tcagatatca agccatttta 34620
aaacaagtat catgggaatt aaaatttaaa attaggcatg tgtggtaaga agggtgaggc 34680
cacattcatc tttaaaatat caggaggtct tagtcatttc ctttcctcta tggttttctg 34740
aaccatgtaa actgaaatta gattgttcaa tcattttaaa tatctaaaag atatagagag 34800
ggttcctttg ttttaacaaa aaaataccaa aggagaaaat acactcataa cgtgaattaa 34860
aatatagttt ttatttttctt ctttaatgtt tattaatgtt tccccgtaaa gtagaacttt 34920
ccttttctgt tctgctttct gtgttctcta cacgtatgtg aaccacaaga ctgaagaaat 34980
ctagcctctg aaaaacccat agttggagtt tattttctct aaaacaattt agttattttt 35040
gtttgttagg ttcaaggatt taagagaccc tttatattat cttagtctta aaccaatttg 35100
ttcccagttt gcgtgtggat gacagctcat caaaactggt caactaggac tgaggctaag 35160
attcaggaag gtaatttaga ctggaaatgg taattcaagg gttgaagaac tcagagctgg 35220
cagagatcac aacagagtcc aagagtttgt agtattcaaa gagaagccag aaagccagga 35280
ccaaagccaa ggacagtagc acttagctta aagtctgcaa agcagagaat catgatgggg 35340
agaatcatga cgttggggag atgttagaat aatttacaaa tattggcatt tacaaagctc 35400
ttagttatct catgattata tcaactcttt aaagtgggta ctgttatctg ctttattcaa 35460
agacattaag agacttggtc caagcaatca cagtaagtag ccgagctgag actgaaatct 35520
tataatttaa ttcttcagac tgaaaatctc atcttctgag atagtagtgt ttttcagact 35580
gctctgttga ttctgaggat tctgcaaagg taaccaaggt gagggcttgg ggatggggag 35640
cctcatccct tctaggccac actccatccc agcttcattc catcagagtt gctcagtttt 35700
aaatattatt ataggccagg caccatggct caaccctgtg gccccactac tcaggaggct 35760
gaggttggag gattgcttga gcccaggggg ttgagtttac agtaagttgt gtttgtgcca 35820
ctgcacttaa gcctgggcaa cagagggaga ccctatctca aaaaaaggtg ggggttatac 35880
acacgattcc tttgaattgt tttgttgtgt aaaaaatcag ttggattagg agcagcagaa 35940
ttctaatttc agtattgtca attagtgact gtaggcaagc ttggtcctat gaaatgggaa 36000
agttagtctt agataacttt tagatttatt tttgccatag tccagtccat ttaaaaagtt 36060
ttaaaaatta tttttagaga tgaggtcttg ctatgttgtc caggctggtc tcttatcatg 36120
tggaaatagt tcatttaaca ttccatgcca gttcagtgaa gtctgtgttt atatcctttа 36180
ggcagaatga gaatgttttg gtgtgtgtat gttttattct agtagctgaa gttaaggcag 36240
ggtgaaacac agtcatacag ccaaacacta gcccagcttc ctacccctcc cagggctccc 36300
tttcccttct atctctgacc aaaaaataag ccctcgaaag gtttttaaaa aataatcttt 36360
ttcctctata ttttaatatt ttaggaacaa gttccagttc cagtctacca cccaacacct 36420
agccagactc ggctagcaac tcagctgact gaagaggaac aaattaggat agctcaaaga 36480
ataggtctta tacaacatct gcctaaagga gtttatgacc ctggaagaga tggatcagaa 36540
aaaaagatcc gggagtaagt tttaattaat ttgtacattt caaagtttta ttttcagatt 36600
gcttttgaaa aatttgcctt ttctcctgca ctttagaaag tcagacattt aatatattga 36660
aacacttcgc ttaatgcaag agaatgttta attttttttc atagctctat tactggctac 36720
tagcaagaga atgttatgta gtaattaaga attttcacag gccgagcaca gtggcttatg 36780
cctgtaatcc cagcactttg ggaggctgag gtcgggggat cacttgaggc caggagttcg 36840
agaccaacct ggccaacatg gtgaaacctc atctctactg aaaatacgaa aattagccag 36900
gcatggtggc acgcgcctgt aatcccaacc acttgggagg ctgaggtggg agaatcactt 36960
gaaactctgt ctcaaaaagt aaaataaagt aaaataagag aaagttagtg ttaatgaaga 37020
gagctttaac ttacagcaac ccagtaagta ggctgaagtt agtgtttctg tttttacagt 37080
tattaaaatg aaaattatct tactctagga ttactctaaa tagcacaagg aaacaaagtt 37140
gcccgaagtt agtgtttctg ttttacagtt actaaaatga aaattgtctt agtccaggat 37200
tacactaaat agtgcaagaa aacaaagttg atctaactga tgcttttcta cctaaaataa 37260
tcccattctc cttcaggtat ctgctataat atgtagctgt ttacaagcac agcagtagat 37320
ttgatgcata tgtaatccat ttctctttgc tctgttgatt gctcctaagt cctgcaaatt 37380
ctcaggggat tgaaataaga gactgtggac agatcaacat taggaaaaaa ttcagtttgc 37440
atgctttggc aataagatga caggcaaagc attctgatgt gttactcaga ttaacactgt 37500
ctttaatggg atttaaagag ttttctgaat aggacacttc tgaaaatgcc ttctgatttc 37560
tctccattgc tctcaacagg tgtgtgatct gtatgatgga ctttgtttat ggggacccaa 37620
ttcgatttct gccgtgcatg cacatctatc acctggactg tatagatgac tggttgatga 37680
gatccttcac gtgcccctcc tgcatggagc cagttgatgc agcactgctt tcatcctatg 37740
agactaattg agccagggtc tcttatctga cttcaagtga accaccattt tggtggtttt 37800
gatcttttgt cactgagccc aaagagccag ggattaggaa ttaagatcgt gcacaaaagt 37860
ttccttaaaa ttcctggatg gctgcagatg ttgggggaaa aagtacgtga tattttagaa 37920
acttagtggg aaaagtagga tggtattttt atgtaaagcc ttgacccaat gtttaaaaat 37980
```

FIG. 3L

```
ataattgtat ttagatcttg ttattgctcc agtacatagg aattgtgtaa agtgttaaca 38040
gcagctgtat ttgtttaaat tgtgtgtatt gaagattagg aaaaagatag tagttatttt 38100
tcctaaatga aataactttc ttctcttccc cttccccacc cgaattcttt tctgaagttg 38160
ctggcatttg ggtcaaggtt ttattaaaag ctacatttta taacactggc acacacaaaa 38220
aagtagtttt aagcttgttt gcacagttct ttttttccat tggaaatgga attcattgcc 38280
ttaggtcttt ttaaatagtg tattattatc gttggggctg gctctatgct tgaaaaccag 38340
tttatttata acctgttata agtgctatat tctgtttgca gttaggaaat gcagaattca 38400
aagtgatctc ctagcttgta agcaaactga gatgcactat ccctttttcta taaaaaataa 38460
gttaatgtgt caagaaacca actctattaa ggtggggttt aatattaccc tttcctatgt 38520
gttttatcta attatttgg ttgttaatat ggtgataatg gaaagtcaag ttaaatttta 38580
aatattaaga attctgattt attgagattg aattatgcca ccacgtttat gtaaaaatga 38640
aggtggcacc gtggtgagac ctaatgagaa atagttactc agttgtaaaa attttgattt 38700
attctctttc ttctgacctc cttgcctctt gtcttgaacc atagcaaaag gatactgcat 38760
ctctcattac tgtagtgctg aggttattga agttatacaa aacacatctc agtctctgtt 38820
tcttggaaag gtatctatta catcctgcta gctgactgac aaaactaagc agggagaata 38880
aagataattg tattttatgt tttgcacaca aacgcagaat ttgtataacc atatgacttc 38940
atagttgtga tctcaaaaaa gaaggaattt ctcctttgtt tcttgcagtt aatgtaagaa 39000
tactttaaat ctctaagctt ctgaagtgtt agaggtagag atggtctagt aaagatgtag 39060
tagtaatgtt ttatccattt agcatgtgtt tatttttttca tatgtactca aaggtgactt 39120
attggttcac ctcagtgata ttacagctaa aaaaatcatt cattagcaaa aggaaaagtg 39180
gtctcaacct aacatcagaa gtgtttctta ttattatttt atattgagtt gaatattgaa 39240
ctctaacagt tttctacata caaaacacag tgtcatgaag gttattcata attgcattat 39300
agaggaatgt agtatgtcat aagtactttg taaagatttg acattcaact gtagtatcca 39360
tatgttgctt aaatttcctt atgagcccca tgatggaaag acttaaagat gaatttgaga 39420
aaaattgaaa gaaattagat tatcaggttc tgttaaattg ttacatgtat cttgcttaaa 39480
tttctgttta ttaatttata tccacccaag tacataaagc aaatttggag gaaacaactg 39540
aagttgtgca atattttctg ataattgctt tttttattct tgtgtttttct acttaaacat 39600
aatgtctgtg tcatcaagta ttatagtcag acttttcttt ttttctagat tgttaaaatt 39660
ggcaaatgaa ctttttttaaa aatcatcttc catgttgcag ttagtctttc ttttcattac 39720
aagtctttca cagaagtttg gtggtaatat tgaaagaact agcattgggc agaatgtgtc 39780
ttttt  SEQ ID No. 1                                                 39785
```

RNF-11 Coding region

```
Met Gly Asn Cys Leu Lys Ser Pro Thr Ser Asp Asp Ile Ser Leu Leu
 1               5                  10                  15

His Glu Ser Gln Ser Asp Arg Ala Ser Phe Gly Glu Gly Thr Glu Pro
            20                  25                  30
Asp Gln Glu Pro Pro Pro Pro Tyr Gln Glu Gln Val Pro Val Pro Val
        35                  40                  45
Tyr His Pro Thr Pro Ser Gln Thr Arg Leu Ala Thr Gln Leu Thr Glu
    50                  55                  60
Glu Glu Gln Ile Arg Ile Ala Gln Arg Ile Gly Leu Ile Gln His Leu
65                  70                  75                  80
Pro Lys Gly Val Tyr Asp Pro Gly Arg Asp Gly Ser Glu Lys Lys Ile
                85                  90                  95
Arg Glu Cys Val Ile Cys Met Met Asp Phe Val Tyr Gly Asp Pro Ile
            100                 105                 110
Arg Phe Leu Pro Cys Met His Ile Tyr His Leu Asp Cys Ile Asp Asp
        115                 120                 125
Trp Leu Met Arg Ser Phe Thr Cys Pro Ser Cys Met Glu Pro Val Asp
    130                 135                 140
Ala Ala Leu Leu Ser Ser Tyr Glu Thr Asn   SEQ ID NO. 2
```

SEQUENCE LISTING

<110> deCODE genetics efh.
Hicks, Andrew A.

<120> A Susceptibility Gene for Late-Onset Idiopathic Parkinson's Disease

<130> 2345.2025002

<150> 60/363,220
<151> 2002-03-08

<160> 2

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 39785
<212> DNA
<213> Homo sapiens

<220>
<221> exon
<222> (3059)...(3308)

<221> exon
<222> (36385)...(36554)

<221> exon
<222> (37580)...(39685)

<400> 1
cccagctatt caggaggctg aggcaggaga attgcttgaa ccggggaagt agaggttgcg 60
gtaagccaag atcgcaccat tgcactctag cctgggcaac acaagcaaaa ctctgtctca 120
aaaaataaaa attaaaaaaa attaaaacta actaaataaa tctatgtctg gccaggcata 180
cagtggctca cacctgtaat tccagcactt tgggaggctg aggcgggagg atcgcttgcg 240
cccaggagtt tgagaccagc ctgggcaaca tggcaagact tcgtctctac caaaaacaaa 300
aacaaatgca aggcatggtg gtgtgcatct gtagttctag ttactggaga ggcctgaagc 360
aggaggatca cctgagccca ggaggttgaa gctgcagtga gcctcgatca agccagcctg 420
ggtgacagag tgagaccctg tctccaaaac aaacaaacaa tgtctacaaa ttctagcact 480
gtgtaaatat tatgtcaaat aattgtataa attattatta ataaaaattt ctttaggttt 540
gacaataatt tgtggggtaa attttctcac tttgtaaaaa ttccagggta cctctgattt 600
ctggtaaatt tagccaatca taaattaaat gagttcttgg aagctcaaca attccaaaac 660
caaaattgtg agtcagtttg ttctttcttt ccttctttcc tttctttttc tttttctttc 720
tttctttctt tctttctttt ctttctttct ttctttttct tctcttttct ctttctcttt 780
ctctcccttc cttctttcct tccttcctgc ctcccttctc tctctctctc tttctccttc 840
ctttttttct tgagactgag tcttactcac tccctcgccc acgctagagt gcagtggcac 900
aatctgggct cactgcgctg caacttccac ctcccaggtt caagcggttt tcctgcctca 960
ccctcccgag tagctgggac tacaggcgtg caccagcaca cctggctagt ttttgtattt 1020
ttagtagaga tggggtttca ccatgttggc caggctactc tcgaatccct gacctcaagt 1080
gatccacccg ccttgacctc ccaaattgct gggattacag acgtaagcca ccacgcctgg 1140
cccaaaatta tgagtctttt caaaaatttt atatcaaaaa ggtgatattt aatgtggaat 1200
ttgagtgcat ttaaataaga tagtgacttc gtagaatagg gactgcagag taagagtgtt 1260
cagacccatt cgagtttaat atagccctac ctctaagaat cctttctgtt cctcactcaa 1320
atgaactaat tgctgtcttt cagttcctct aaaaaccttg gggtccgtgg cagggtgtgg 1380
gcttaggact cacctgatta cttgtcttcc caggtggtct ctgggctcta tcaggaacaa 1440
gccaggtatg atcatccttt tccccagtcc ttagcgtggg gtctggtata cacacgcaca 1500
cacaacttca gtaaatggtg aatgaaacga aaggcagtca agatccacct gttgaatgaa 1560
cgtctatcag gagcactgtc atccttcttg cttccttctt ttgcatatgt tgtttgagct 1620
gctcattcaa ccaaggagag agcatactct accttgggca gtgtgccagc cttgcacaga 1680

gggacagatg attcagcttg atcaggctgt gacagagaaa agctcagggt ctgagagcac 1740
acaggatgcg actaacccag actggagggt caggagagtc tccccagggg aaatgaagtt 1800
gcatccagaa ggattcctga gaataatgat tagcctgatg agaagcagaa aggttgggag 1860
tggagggaag tgggtttctg tgagagaaac aatatggaca aaggccccag gcagaggcaa 1920
tagtattcca gataacgaaa agcattttag agagtctagt ttagaatttg tgaaggaggt 1980
gacaatgctg ctaatgagat tgaaagggag tcagatttgt aaaccattta aggattgtct 2040
taaggacagc cttgaggaga cactgaaggt tttaacagga gaatgacatg atccgattta 2100
aattttaaaa agataactct gtcaactgtg ccaacagtga attggatgtg ggcaatgatg 2160
aactcaggaa cagctgagag tctattgcag aagccaaggt gagagatgat ggtggcttgg 2220
actttggtgc tgggactgga aagaaatgaa caatctgata tacaactgga agaataatct 2280
ctctggctgt gatgtggaag gataaggacg tgattggaca ggcagggatg gttcagtaac 2340
gagctaggac tttattgtga agggataaag gggaagcttt taacatttta atcatgggac 2400
tgatccattc cgatatgggt ttatataaat agaccgcttg gcggagggtg gtgagaagac 2460
tgtcaggcag tgcacaggta agaactgagc cctagacggg gcaggagcaa tgaatgtggg 2520
gaggcgcggg agacgggaga ttggtttcgg agccaggtac ggagccaagc agtcagagag 2580
aagcaaaaat tgacgcagca ccgagaacag ctgctagtgg agaggggat gggcgccgcc 2640
ggagccgccc tcgcctcctg gcgccggggg cgggaccaga gcacgcacgg cgcgatgacg 2700
cagtgcccga gccggcctcg ttctggctgc cgccgtctcg ctgaggcgga gtaggatgag 2760
gcccgcggcc gtggctccgg cgtcggcggg ggcagcagca gctgaggcag cagctgaggc 2820
agccgcgacg gccgcgcccc cccccgctga cctggattag gcccagcagc tccgtggccg 2880
ccgccgcccc gcgggggggc ggggtgggga agtgcttcgt ctccccgccc ccccccccc 2940
atcgctgcct cgcaggaggc gaacgccgcg accccagcgg agcccgcggc ccagccttga 3000
tcccccaacc ccgggggctg gcatgagcgg cccctcggcg gcaccgtggg gcggtggagt 3060
cgcctccgcc tgatccccgg cctgtcgccc gaccccacct cgccaaccga ggcggaccgc 3120
ggagtgtgcg aacgacccca ccgctgcttt ctcctccccc agatcacgca ccccagctcc 3180
ggaagatggg gaactgcctc aaatccccca cctcggatga catctccctg cttcacgagt 3240
ctcagtccga ccgggctagc tttggcgagg ggacggagcc ggatcaggag ccgccgccgc 3300
catatcaggt aggggagggt gtgtgttggg gggagcgagt agaggcagct ctggcggaga 3360
ttgagggggc ttcggggccg tctctgcccc tggcttcggc aaggcctggc ctgggcggca 3420
ttgacccctt agggctggat ctgagggcat ttgctctctg atctcagagt aaggttgatt 3480
cttcgcctgc cctcgggcac cgtgtgggga tgataccctc tgcccttgag ttctgggatc 3540
agagtatgtg atcacatccc cttggccagt gtctggtggt tcgcttgtgg agtttgggag 3600
tgacattttc ctgtctttat gtctgaatcg tatgatttcc tcatcgcacc gccaactgga 3660
ttatgggttc tggataggtg gttgccaagc accacccccc cgccccattt atgttttcat 3720
gtgtggtcgt cctcttaccc acctccctcg tctttttgtt aattggatct ggcaggtttt 3780
ccctctgctt gggagacgga tctccaagtg gtgattttaa ccctagcgac atcgagttac 3840
ttgtgtgtgt gtgtatatat atatatatgt gtatatatat atatatatat atgtatataa 3900
atttcaatcc ctcctctttt tgctccaccc atatgacaca ccttttaaac gtgctgactt 3960
aagaaaactt gattagcttg aactttggaa acctttaaac ctggcttggg aaattccttc 4020
ggttactcgt gttctccctt tgtatcaaag actctcctaa atacgttgta ctgctttttg 4080
tacaataatc agtagtccca gatgtgttgc agaattcgag tgaagacctc ttaaactcaa 4140
aataggtatg tggatgttat aatccacaat agatgtgtga aaacctttcc tgtacttttt 4200
ccattttaga tgatattgaa tacacccttt tccagagggc ttcaatagta gtagtgctta 4260
atttcttcct gaagagttta gttttctgac tgggtaggct cctttgttgt ttcatgtgat 4320
gacatcatta tctagtgttt tctttttctca ttcctgtgta aaccaatgtc tcatcacacc 4380
agaaaaacct gatcatatat atgctcgttt tctgaataaa attcttctgg gggcatttca 4440
tatataccag taaggtccag tcttgattgt ggtattacag aagcctccat gacttctccc 4500
tcttacgttc tatattccag gcctaccaaa tacaccaggc tgttcacagc tttgttttgc 4560
tccagttcct ctctgcttgg agcccttcgc ccgggtcagc caggtggaca tctaatcctc 4620
tttgagaact cagctttcat gtctgaagtc ccagcattcc cttccttccc agcccgaatc 4680
ccttgtgtcc ccactgaaac ctctacactt ccataagcaa tatctataac actttagtgt 4740
tgtatctttg ttagatgtgt gtctcatcct tgtaatcgcc tccttcagtg caagctgcgt 4800
gaaggtagag aatctgtgtt tccggtttgc tgttgtatcc ctcgcatctg cctcagtagg 4860
gactcagtaa ctgggagata aattacgagt aaaagccctt gagaacaggg actctccccc 4920
cgctcagttt ctggcacata agcacctaaa tgctgaatta attaatgaag ttgaatttag 4980
tattttctgt atttacacaa atatttttttc caaagtaagt tgtcacaggg tagtctctta 5040
aaaatcaaag ctgaatctgg gtgtctttac aagtaccttt gagtgaagca agcaagctat 5100
gtttatcctt cactgtcttt ccctcctatg tttactatac aacaaaagtg cttatttaac 5160
ttttttttt taatgagatg gagtctcgct ctgtcacccc ggctggagtg cagtggcaca 5220
atctcactgc ttactgcaac ctccacctcc cgggttctag cgattctcct gcctcagcct 5280
cctgagtagc tgggattaca ggcgcgtgcc accacaccta gctaattttt gtatttttag 5340
taaagacggg gtttcaccat gttggtcagg ctggtcttga actcctgacc tcaagcagtc 5400

cacccacctc agtctcccaa agtgctagga ttacaggcat gagccaccgc atctggcatg 5460
aagtttttat tttaaaatct cttatgctcc ctaccctcac ccccttctat ttttaaaaaa 5520
atagagacag ggactcctat gttgcctagg ctggtcttga actcctggcc tcaagcaaat 5580
tctcctgcct cctatccccc ttcttaattg aatatctttg acttgttctc tttctcctat 5640
ccttatgata attgtactat taaaggtact agaatttatg gaaactgtgt ttgccttttc 5700
cagaagaact aggtgatgat gagggatgag gggtttttgg tgtcatgaag gaacttttt 5760
tggaaatgtt atatttttga cagtgggagt tttttatggt tgcagtttcc tgaagtgaat 5820
gacgcagact aatatacaaa ttcagtatca atcagaagaa aattacttta aaacatggga 5880
gcataaaaag aaacttttgg agaaaaagta ttgtccccca cccccaaaga aataagatct 5940
gtcttaggaa tggttatttt agatgatctg agaaacagca atccaaggtt atttattcat 6000
ttatttgagt tgagaatctc cctctgtcat ccaggttaga ctggattgct cacttggcag 6060
cctcaaactc ctggcttcag gtgatcctcc tccctcagcc tcccaaagac tgggattaca 6120
ggcgttagct acctcacaca gcagatttat agtattctca gttgataaaa gtaacttttt 6180
caacaaatta tatgtaatta acagtgaata gctgattgta aaggaacatg ttctcttgag 6240
gtgagtagtt aaaagatgtc attgtttgct gctttatatg tacaatgtac atataatgac 6300
aagagccagt tacagttatt tcctattaac atctgacaaa tttgttttta tttatttttg 6360
ccttaaagta aatccctgca ccattatgat ccaagtttgt aggaagtgta tatttctttc 6420
catggtttct tttgtagcct cctaacagtt atccctgctt tcacccttgt cctcgctaag 6480
gcctgtcctc aatacaacag ccacagttat tttggtaaaa ctgagcatga cagtctcccc 6540
aaccctctgt gtggatcctt gtttcattca aagtaaaaat cagattcctt atagtgtcgc 6600
ataaacccta cataatctgt acccccttgtt tccacgcatg tcccttgaca cctacagcta 6660
acccctttgt tggcttccct ccagcttccc tgtcctcatt gctattctag gacctttgtc 6720
ctccttattc tggaatactc ttccccagat ttcatggctg actgactccc ttacctcctt 6780
caagtctttt ctcactgccc ccttaaacct tctctgacca ccctgtttga aagtgcaacc 6840
tgccctaata acctggaagt ccctctcacc ctgttgtttt ccccccaat accatttatt 6900
acattccaac atgtgttaga aagttgcgtt gtctattacc tgtcttctcc cactaggtag 6960
aaaggaggaa aggggatttt tttttttaat tggtgtttcc taagtactta taacagtgcc 7020
tggcacatgt taggcattca gtaatttggt tgaacgaatg aatagatttc ttttaaatga 7080
actaatgtat ttgaggcagg tatagttgca ctgcttacaa gaactctgca tattcctagt 7140
gccatttttg cttctatatt gtagatcagg ttttttttttg tttgtttttt tgttttgtga 7200
ctgggcctcc ctgtgttacc caggctggtc tctctctgta tatatatgtt tgggtttttt 7260
tgttttctct tttttgagac caagtctcac cttgttgccc agctagagtg cagtggtgtc 7320
atcttggctc tgcctcttgg gttcaagcga ttctcctgcc tcagcctctg gagtagttgg 7380
tactacaggc atgcgccacc aaggccaact gatttttgta cttttggtag agatggagtt 7440
ttgccatgtt gaccaggctg gtctcaaact cctgacttca ggtgatctgc ccaccttggc 7500
cttccagagt gctgggatta caggcgtaag cccctgagcc tgccctttat atttttaaa 7560
tttaaaaaat agagacaggg tctcattatg ttgctagtct caaactcctg ggctcaaggg 7620
atactcctgc ctcagcctct tgagtagcta gggttacagg tgcacttcta ctctgcctgg 7680
ctcaggtttt aaaaaaaatt tatttccttt agagatgagg tcttgctgtg ttgcccaggc 7740
tggagggcag tagctattca caggtgcaat tatagtgcat tgcagcctcc acttcctggt 7800
ctcaaaggat cttgtggctt cagcctcctg agtagctggg actacagggt tacaccaccg 7860
tgtctagctg ttgtagatca gttttaagga atcctaaaaa gtcaaattcc agtaatatta 7920
aatacagcaa ttggtttaca ctaagtgtta agtgatttag gggtttttac cgtgtatata 7980
tatcaccgca agatatttat gtaatgggaa agtacttggg ggagaaaaag aaagtacttg 8040
aatccctatt cttgaagtat tggatgcaaa gagatttagg gtgtttgtca aattagaaag 8100
tagaaacaaa tttgtcgcac cctctaccca actcccttat tcctaaatgt ttaaaaataa 8160
tacagaggaa atctgcccag cgaaggcaat cttggaagga cagtgagtgg gaggataagc 8220
caggaaaggc tgaagaaggg aggtccatga gaggggggagg ccaggtaggc agcccagccc 8280
tgaaaggcat atgtaaaaca gcttgaggta cccctgtgtc actgtcaaga gctttgttat 8340
gtgctaaatt gagggaggca ctttgttaaa agagtgtaat attcattgaa caacaaacca 8400
tgttacattt tctaatctga atatataaaa ataaggatag gtgtctctca attttgcagt 8460
agttatatag tttgagattc gccagtcttc ttccccgacc tatacttttt gattgataca 8520
tttgaaattt taaaatcctg agaaaataag tttgtatagc attgaatgca agtatagctc 8580
tgcttttaga gtcagattgt cctgtgttca aattctggct ctgccgctta catgtaaaat 8640
atgaaaaagg aaagaattta tcagattgtt ttggggagct tttttgacaa ataagaaaat 8700
gtaacatttt ggatatttca aattgtgtct gcataaagct ttcttcaaga tacctttttt 8760
ttttttttt tttggctggg cacagtggct cacacctgtc atcccagcac ttaggtggca 8820
gaggcaggag gatcacttga ccccaggaat tcgagaccag cctgggcaac aaagtgagac 8880
ccccatctct attaaaatac caaacaaaaa caaaaaacaa acaaaaccaa aaaccctctt 8940
ttgttgttaa aaggattggt ttatcctgta gaatgcacta aaaataaccc ccttccttat 9000
aaaaaatcta gaatgtactg cttatttttа agcatgcaat gctataatgt tttgtggggg 9060
atgggcttgt ttttttttaga gatggagtct cactttgccc aggctagttt cgaactcctg 9120

ggctcaagtg atcctcccac catgtctccc agtgttggga ttacggatgt gagtcactgt 9180
gccctgcctg taatggtata ttaatgcctc attttcctga taacttacga agttaaattt 9240
agataattaa agggttgtgt ttttttatta gcattttgtt taacatttgg gtggaaaata 9300
tctccaaatt tatttcacgt atctaggcct tattatacaa ttattcagct tttaacaatt 9360
taccttttg aagttttatc cctccccttc tttatgctca agaagttcag ccttcttatt 9420
gaggggcaga tcaactaact ccacctttgt taataaagcc ttcactgacc attttttcc 9480
tgtctctgct gatttgtcat gtacagtttt tttgtgtttg taacatttat ctggaaatca 9540
tatattgcct tattagcttt tgttgcttta acagttgttt ctccacatca gtttcctcaa 9600
gaatgagaat gatacattaa acttgtaaga tatcttagtg gagtagaatg ttaccttcca 9660
tattgtagac acccagtgaa tatttgatga tagactaaat tagaactgtg acttcttacc 9720
ttgtaaggtc accatcataa ttgtcctcat cagttcttac agcaacataa ttgatgtctc 9780
ctcctcattc attttatacc atgttgccag ataaatttca tcccggatca gatcacttgc 9840
ctcataatcc ttaatgattg ttattattat tattattatt ttttgaaaca gagtcttcac 9900
tctgttgccc aggctggagt gcagtggtgc cacctcagct cactgcagcc tccgcctccc 9960
gggtccgagc aattctcctg cctcagcctc ccgagtagct gggattacag gcatgcgcca 10020
ccacacccag ctaatttttg tatttttaat agaaatgggg tttcaccata ttgctcaggc 10080
tggtgtgcaa ctcctaacct caggtaatcc actcaccttg gcctccgtaa gagctgggat 10140
tacaggcgtg agccaccacg cccggcccтt aatgattctt tattaggact aaccactgaa 10200
gatatctgac ttagcattca aagtcatgat ttggctctac atatactttt tcagctcttc 10260
atagccttgt gctttctttt atacaacctg tatactccag tcatactgag cagtttgctg 10320
tattccatgt atgtcttctg ctttcctgtc tttgcccatt ctgtatccat tgcccagaat 10380
gttcttcctt aatccttccc agtctcattt ctaatatcca gggcccagtt caaatgccaa 10440
gaatccacga attctgattt ctcctacata cacttttctt aatctgaact cctttattca 10500
tgcaccattc ttaagacact ttacccтttt aattttaatt atttgtgaca gtgtcacatc 10560
tgctctttta gaattcaagt ctggttcatt ttcacattgc tagaactaca attttcacaa 10620
agcacgcact tatactttag tgaatgtaat gatcagcatt ttgctgtaag agtgtgcaat 10680
tgtatggttg ttagtgtcta ttaagcattc ccagaatcct gtgctttagt gtttggccta 10740
tcacttctaa tttataaagt acgttgagct ttgggaaagt taatttcatt atatattttt 10800
tgaatttgtt tttttttttg tttgtttgtt tgagacgaag tctcgctgtg tcacccaggc 10860
tggagtgcag tggcgcgatc tcagctcact gcaagctccg cctcccgggt tcacgccatt 10920
ctcctgcctc agcctgtccc gagtagctgg gactacaggt gcccgccacc acacctggct 10980
aatttttttgt atttttagta gggacggggt ttcaccgtgg tcttgatctc ctgaccttgt 11040
gatcctcctg cctcggcctc ccaaagtgct gggattacaa gcgtgagcca ccgcgcccag 11100
cctgaatttg gttttgtgtt ccatgaagtt cagcttgcct aggtgacttt tataaataag 11160
taaccagtta tttgttattt gtgtgacccc attgttgcag gatttttctt tgtttctgcc 11220
tgacctctca ttctgaatgt gctttaataa aggctgattt tttggcatta agattgtatg 11280
catcactaag agtaagatag tctgttcaga gtttaattct agagctaatt atagggtaac 11340
ttgagaagat ggagatggga tggtgagtta cttggtacat: aagatctgag cacttgattt 11400
tcttaaaata ttttttaggt aatagtagaa acctttgttg cagaattttg atctttaaaa 11460
gttttaatag aaatacttat gtagattggg attctttttc tttctttgag acataatctt 11520
actctgttgc ctaggctgaa tgtagtggca tggtcacggc tcactgaagc ctcaacctca 11580
cgggcccaag ccatcctccc acctcagcct cctgagtagc tgggactata ggcgtgcacc 11640
actttgcctg gctaattttt atttttttgta gagacgggtc aagcagtcct cccatctcag 11700
tctcctgaat agctgggacc acaggtgcat accaccacac ccagctattt tttgtatttt 11760
tagtagagac ggggtctcac tatattgccc aggctggtct caaactcctg aactcaagca 11820
gtccacccac ctcggcctcc caaagtgctg agattacaag tgtgagccac catgcccacc 11880
caacagcctt ttttttgga gatggagtct tactccatcg cccaggctgg agtgcagtgg 11940
tgcagtctca gcccactgca acttctgcca cctgggttca agcgattctc atgcctcagc 12000
ctcctgagta gctgggatta gaggcaagca ccaccacacc tggctaattt tttgtatttt 12060
tagtagagat ggggtttcac catgttggcc agactggtct taactcctga cctcgtgatc 12120
cacccaccta ggcctaccaa agtgctgaga ttacaggcgt gagctactgg gctggacacc 12180
aacagcctat ttttaataac ccatatgagt acaggacaga aatacatccc taactctgta 12240
cctgaagtaa ggaaggtaat acataaaaga catatttgca aacgacttga ggcaagaaaa 12300
gaaatacaac acatctctaa ttctgattgt atctgcaatc aacctcttat cacttccctc 12360
ttcatgtgca gaatttacct tggctgagac acagtcaaga aagaagacag gtggccaggc 12420
acagtggctc acacctgtaa tctcagtact ttgggaggct gaggtgggcg ggtcacctga 12480
ggttaggagt ttgagaccag cctgaccaac atagagaaac cccgtctcta ctaaaaatgc 12540
aaaaatagcc gggtgtggtg gcacatgcct gtaatcccag ctacttggga ggctgaggca 12600
ggagaatctc ttgaacccag gaggtggagg ttgtgttgag gctaggttgc gctgttgcac 12660
tctagcctgg gcaacaagag tgaaactcca tctcaacaac aacaacaaaa aagaagacga 12720
caggctgtgt gcagtggctc atacctgtac accagcactt tggaaggcca agtgggaagg 12780
attgcttgag cccaggagtt cgaaaccagc ctgggcaaaa tagcgagatc tcgtctctac 12840

aaaaaaataa aaaattagtc aggtgtgttg gcacatgcct gtagtcccag ctataggctg 12900
aggcagaagg aatgcttgag ctcaggagtt ggaaggtata atgagctgtg atggcaccat 12960
tgcactccag cctgggcaac agagtgaatg ggggcatggg gaaggccagg tgtggtggct 13020
cacacctgta atctcagcgc tttggtaggc tgaggcggga ggattgcttg catccaggag 13080
tttgagacca gtctaggcaa cacagtgaga ctcccatttc tacaagaaaa tatttaggct 13140
gtaagtggag agttattgtt tagtgggtag agttttctgt tttataagat gaaaagttat 13200
ggagatggat agtttaatgg ctacataact ttatgtaata ccactgaact atacactgaa 13260
aaattggagc tgggcttgtg cctgtggtcc cagttaatca ggaagctgag gcaggagggt 13320
tacctgagcc caggagttcg aggctgcagt gagttataat tgagccactg cattccagcc 13380
taggcagcag agtgagacct tatctcttaa aaaaaaaaaa aaaaaggaaa aaataatagg 13440
ctgtaaacag ttgattcctt accagtaact tttaatattg aattaacaaa tatttatggt 13500
atttttaaca tgctcagagt tcaggctttt ttcagctgtg atctaaatga aagtgagagt 13560
gaggtggtga gagaagtcag tggaaatgtg agtgaaagat tgaggaagag ctggcttaga 13620
tagccacagt gtgtagttat ttttttttt agagacgggg tttcaccttg ttgcctaggc 13680
tggagtgcag tggctattca caggtatgat catagcacac tacagccgca aactcttgag 13740
ctcaagccaa cctcctggtt cagcctccca agtagctggg accacaggct catgccacca 13800
catccagtca aagtgtctgg gttttttggt tttcaattag gttcctcagg ttaaatggaa 13860
gtatgtggtt ttttgttttt ttgtttttt tttttgtaaa gacagtgtct cattatgttg 13920
gccaggctgg tcttgaactc ctggccccaa gccatcctcc tgcctcaccc tcctaaagtg 13980
aagtgctgga ataacaggca tgagctgcta ggcctggcca aagttgaaaa gtaaagtgat 14040
aagggagaag taaatgtatt ttatcctctt ggagaatttg cttgacctat tgatttttat 14100
attcaaattg gccttttgaa ctcttagatg agttccttat tttgacaatg aatgcgtgtg 14160
tccctatctt ttaattaaat tgccagccta gagtgtactc ccaggggcca aatgagtaat 14220
tatcttgtta cttaatctag aaagcatact gtaaggtaac taaaggatct gaaagtgtgg 14280
gaatgaagaa agcaaccctc caaatatatt tgcttctaaa actcttaaaa ttttctgaac 14340
actcataaag ttgaatggct ttaagaagag tgagacttta taattgaccc agttagttaa 14400
agttacatgt aattttgtat aacagtttaa accttatttt tgtggaagaa aaagaagcag 14460
tttaccataa ttgagtagcg aggccttaag ggtaagaaat accatgcttg atacctagtt 14520
tcttcttttt ttttttttt ttttttgag atggagcttc gctcttgttg ctcaggttgg 14580
agtgcagtgt cacgagaaaa tacatttaca gtgctatact gtattttctt tttctttttt 14640
tttttttgg ctttgagatg gagtcttgct ctgtcaccct ggctggagtg cagtggcacg 14700
atcttggctc attgcaacct ctgcctcctg ggttgaagcg attctcctgc ctcagcctcc 14760
cgggtagctg ggattatagg cggcgccacc atgcctggct aattttgtat ttttagtaga 14820
ggcagggttt ctccatgtca ggctggtgtc gaactcccga cctcaggtga tccgctcgcc 14880
tcggcctcac aaagtgctgg gattactggc gtcagccacg gcgcccggcc tctagtttct 14940
tctatcagtt agcaagtatg ttttaagctc aacggggaaa ggagaaatga tgagtctctt 15000
acccctagga gcctacagtt aagggataag acatgaagca attatgaggt cagcatgaaa 15060
tagtacataa tcaattatta aatggttttg tatatagttt tgaagttgtg tgatcagttt 15120
gtgatagggc agtgaatcac aaatgtactt tttagcttct gtatcctttg gataatctta 15180
cttggaactt gagtataaaa aatgaataaa aacaacagag ctgccctgat taaattgagt 15240
ggtggagctt ggagtcaaaa agatagcctt ctgtattcac aggttccaca tcctctgatc 15300
catccaatca cagaaaatat ttgggaaaga aaacaataaa aaataataca agagctaata 15360
caaattaaaa aaatatatag tgtaactatt acataacatt gtattaggta ttataagtaa 15420
tctagcgatg atttaaagta tatgggaggg tgtacaggtt atgtgcaaat atgccatttt 15480
atataaggga cttgggcatc tgaggatttt ggtgttcaag gattttggag tcagtcgggt 15540
cggactaatt gttagatggt ggacatgaat ttagaggtca ttgagttgtg taattttgta 15600
cagtaaaagt atgtgtgtgt aaattttaat taaaatgtat tatacagaaa ctagtgtatc 15660
tttaacagaa tatcttgagt ttcttttgat tgactgattg attggtcgat tgattgagat 15720
agggctctcg ctgtgttacc caggatagca ttgggctcaa gtgatcttcc tgcctcagcc 15780
tcccaaagct ctgggattac aggtgtgagc cactgcaccc agcaggtttc ttaattacgt 15840
aggctagtgc acttgcaaaa tatagccaca actataatat agtctgtttc caagccccca 15900
aaaggctccc tcctgccccc tccaacagtt agtacctttt ttccaaaggt aactgttggt 15960
ccttatgacc tcctatcacc catatttgtt ttgtgtggtc atgaatttca aactgaataa 16020
cattctagca ttctttgtga aatttgtgtt cttacgtaat ggcagtagtt cacattcttt 16080
ttcattgcta tgcagtattc aaaattgtgt aaatatacca cattttactt attcattcca 16140
ttatgtgtat ttgggtagta tcaatttttg gctgtataaa taaagatgct gtgaacatta 16200
tacctgagca cctattttga gtatatatcc agggatagag ttactggata atggagccac 16260
atgtatattt agtttttcct gcagcatttt actacagaaa ttttcaaaaa taaacaagag 16320
ttgacagaat tgtttgaata cccatattca ttcctaccac ctagattctg cagtaaacgt 16380
tttgctgtat ttgctttcac acatacctat tcgtcgttct tctgtcaatc tatcctgttg 16440
atgtgtttca aaaataagtt gccaaatagt tttccaagtg gttgcactaa ttttctgctt 16500
gtactgacaa tatgagagtt tcatttgtgc cacgtcttca ccagtatctg gtattgtcaa 16560

tctttatttt tgaaactttt tttagctttg aagtataatt gatataatta tatatactta 16620
tgatgtacac catggtgttt taatatatgt atacatcatg aaatgattac cacagacaag 16680
ctaatttaca gatccttcac ctcacagtta tgatttgcta ttgtgagaat atttaagatg 16740
tattttcata gcagttttca agtatacaca atacattatg ctaagctgta gtcactatgt 16800
tatacaacag atctctggaa ttgtagatct gttttatttt agacatttta ataggtgtgt 16860
agaagtatag aaatgcagtt gacctttgaa caacatagag ttgaatagcg tgggtccact 16920
tacagtgcta atttgtttac acaagtatgt tggaattttt gtttgagatt tgtgacaatt 16980
ttttttttt acgcttaatt cgctttattt tccttgtcta aaaaccctat gttgtagcca 17040
cagctggagc ctgggtccgc tgcacggaga ctctggcgtg ggccttgatg aggtggtcag 17100
tgaattcctg atcgggagac ttggtaaata cagtctcctt ccagaggtcg ggggtcaggt 17160
agctgtaggt cttagaaatg gcatcaaagg tggccttggc gaagttgccc agggtggcag 17220
tgcagcccca ggctgaggtg cagcagtcat cgataccagc catcatgagc agcttcttgg 17280
gcacaggtgc ggagacaatg ccagtgcccc tgggtacagg gatgaggtgc accagtgcag 17340
agccgcagcg gcctgtcacc ttgcaaggga cggtgtgggg cttgccaatc ttgttccccc 17400
agtagcctct gcgcacgggg acaatggaga gcttggccag gatgatggcc ccatggatgg 17460
cggtggccac ctccttggag cacttaacac ccagaccgac gtggccattg tagtccccga 17520
tagcaacaaa cgccttgaac ctggtgcgct ggccagcacg ggtctgcgtc tgcactggca 17580
taatcttcaa aacctcatcc ttgagagagg cccccaggaa aaagtcaatg atctctgatt 17640
ccttaatggg cagggagaag agatagatct cctccaggga cttgatcttc atgtccttca 17700
ccaagcagcc cagcttggtg acgggcatcc actccttatc ctcggccttg cctccgcgag 17760
ctccgcggcc tcggccccgg ccccgtctac ggctgcgacc ccggccccgg atgccactgc 17820
cgaaacctcc gcggaagcca ccgcagttcc ccatcccagg gcctccgggc ccccccccgc 17880
tgcactggtg tcatccacca tttggtgttt tctcggagaa gaaaggattt gtgacaattt 17940
gaataaactt aggtgattct catttttga aatatcaaaa aattaaggta tgtcatgatg 18000
gatgcataaa atatgtaaat actagtccat tttatcattt actaccgtaa agcattcaca 18060
tctattataa aaaattaaaa tttgggctgg gcatggtggc taatacctgt catcccacac 18120
ctttgggagg ccgaggcagg aggatcactt gaactcagga ttttgagaca gcctgggcag 18180
gagagcaaga cctcatcttt acaaaaaatc aaaagaatta gctggagccg ggcgaggtgg 18240
ctcacgcctg taatctcagc actttgggag gccaaggtgg gtggatcact tgaggtcagg 18300
agtttgcgac gagcctggcc aacatggtga aaccctgttt ctactaaaaa tataaaaatt 18360
agctgggtgt ggtagtgcat gcctgtaatc tcagctactc aggaggctga ggcaagagaa 18420
ttgcttgaat ccgggagaca ggtcgcagtg agccaagatc atgccactgc actccagcca 18480
gggtgacaga gcaagactcc atctcaaagc aaaaaaaaaa aaaaaaaaaa agaaacagaa 18540
aaagaaaata cagtataaca ctgtaaatgt attttctctt ccttatgatt ttcctaatat 18600
tcttttctct ggcttatttc gtaagaatgc aatatgttat acatgtaaca tacaaaatat 18660
gggttaatca agtttgtgtt atccataagg tttctttggt caacagtagc ttattagtag 18720
ttaagttttt gggaagtcaa agtttaggcc atgcatgggg catatacccg taatcctagg 18780
tatatgtctt gggaagccaa gacagaagcc aagacagcag atcatttgag ctcaggaatt 18840
cgagaccagc ctgggcaaca cagcaaaact ctgtttctat aaaaaataca aaaattagct 18900
ggatgtggtg gcgggcacct atagtcccag ctacctgaga ggctgaagca ggaaaattgc 18960
ttgagcctgg gaggcggagg ttgcagtatg ccaagattgc gtcattgcac tccagcctgg 19020
gtaacgggag tgaaaccctg tctcaaaaaa aagaaaaatt atatgaaggt ttgcgactgc 19080
actgggggtt gagacccctа accccatgt tcaagggtca actgtatttc attgtggttt 19140
taatttgcat tttcccagtg agtagtagag ttaacctttt caagtgcttt ttggcctttc 19200
gggtattgtc ttttgtgaag tcttttgtcc agtttgtgtg tgtgtgtgtg tgtgtgtgtg 19260
tgtgtttggg ggggacagag agtcttgctc tgttgcccag gctggagtgc agtggcgcga 19320
tcttggctca ctgcaacctc cacctcccag gttcaagcga ttctcctgcc tcagcctccc 19380
aagtagctgg gattataggc gcccaccacg cgcggctaat ttttttgtag ttttagtaga 19440
gatggggttt caccatgtca gccaggcttg tctcaaaact cctgacctcg tgatccgccc 19500
gcctctgcct cccaacgtgc tgggtttaca gatgcgatcc accacgcccg acgttgtcca 19560
gtttttaatt gggttgtctg tctgttaaaa acttgatttg taggagttta aagtatattt 19620
tactaataaa aaagaatggt aattgttctc tgcataatat ttagaaaata aggccgggca 19680
caggggttca agcctgtgat ttcagtgctt tgggatgcca aggtgagagg atcccttaag 19740
ggaaggaggt tgagaccagc ctgggcagaa tactgacacc ccatctctac aaataatata 19800
aaaaagttag gtgtggtggc atgcacctgt agtcccagct acttggaaga ttgaggcagg 19860
aggatcactt gagcccagga gctcgaggta acagtgagat atgatcacac cactacactc 19920
cagcctgggc aacctcaaga tcctgtctct taaaaaaaaa aaaaaaaagg aggtggggga 19980
aacgaaataa agaatagaag gaaagtagag attgtgttag atcctagaat gtgaagataa 20040
gtaaaatgat ctccatgttc tctgtacccg cgttctagag agggatgtag gtaaatggat 20100
ataaaataca ctgccataaa ttttatgaag aagaatgttt agggtgctat aatttgctca 20160
acacttcatt atttttgctt tcttgatatt ttaaataata ctatgaatgt ttaatcgata 20220
aaacacttct tgttttagaa tattttattt gtataaattt ttttaggaat tgctggatct 20280

tgccaggcac aatgactcat gcctgtaatc ctagcacttt gggatgtgga ggcgggagga 20340
ttgcttgagc ctagaaattg gagactagcc tggacaacat agtgaggtcc catctctaaa 20400
aaaaaaaaag ggaaggaatt gctaaatcaa aggatgtgag gattttaaag ctctacacga 20460
tcaatcggat catcctaaat aatttatcga aacagtggaa aaatatgttc cacaataaag 20520
catacctctt tggtgaccaa aatgtggcat tgatgaatat tgggggttc cccctaaaac 20580
tttaaccccc tttgtgttgg cacattcatg agataattct gttgattagg ttatagtttt 20640
ttttttaaat caaagaataa aacatggtct ccttgattag agttagatat tccttgcttc 20700
cttgaatgaa ttccttttc tttttctgtt tttgagatgg agtcttgctg tgttgcccag 20760
gcctgtagta cattggcacg atctccactt actgcaacct ctgcctccca ggttcaagca 20820
gttgtgcctc agcctcccga gtagctggga ctacaggtgt atgccaccat gcttggctaa 20880
tttttttgta tttttaattt ttttttgaga tggagtctcg ctctgtcacc caggttggag 20940
tgcagtggca ggatctcggc tcactgcaag ctccatcccc tgggttcaaa cgattctcct 21000
gtctcagccg cctcccaagt agctaggatt acaggcgtgt gccaccactc cctgctaatt 21060
tttgtatttt tagtagagag gggatttcac tatgttggcc aggcaggtct cgaactccca 21120
gactcaagtg atctgcccac ctcgtcctcc caaaatgctg ggattacagg tataagccac 21180
tgcacccggc catttttttg tatttttgta aaggcagagt ttcgccacat tggccaggct 21240
ggtctggact tcaggttgcc ctggcccctt aaagtgctgg gattacaggc atgagcaccc 21300
agcctagttc ctttttcaaa tgaatagcac ataagtatat gctggctctt actaacttta 21360
ctctgggaga aaaaaacata gctccaattt tcttttcatt ataggctctt tcttgaatag 21420
aatacttgca tagaaaagca atagcttata cattggtgag gggagtgttt acattgcccc 21480
caacatactg tttaatgcat aacatacatt attttattta attatcacat agtatctgta 21540
aagaggaact atattatttt taatactatt aatatgctat ataatatgta tatatatgtt 21600
tctactactt atatatcatg caatgtgtac attatagctg acatagagac tcagcagttt 21660
gtactctgga taagtttgtt acatttaatt ttttatttt ctttttagtg acagggtctc 21720
acactgttgc ccatgcttga gtgcagtggc acaatcatag ctcattgcag cctcaaactc 21780
tgggctcaag cgatccacct gcctcagcct cctgagtagc tgggactaga ggtgcatgcc 21840
accacacctg gctaattttt gtctttttgg cagagacaag gtctcactct ttttcccagg 21900
ctggctttga attcctgggc ttaagcaatt ctaccttggc ttcccaaagt gttgggatta 21960
caggcgtggg cgactgcacc cagctgtgtt tcactttctt gctggtgttt ttggaagcac 22020
aaaagtgatt aatcctgata aaaaccagtt tgtctatttt tttctttggg gagcttttgg 22080
tatgttcaag gaggggggtca aacttgattc ttctgcatgt gactagcctg atgttccagc 22140
accatttgtt gaaaatagtt ttctttttctc attgaattat cttggtattc ttgccaaaga 22200
tcagttgtct ataaatatga gagtacatct ctggactctt ctgttccatt ggtctatatg 22260
tctgtcctta ggctagtact gtactgtagc tttataagta aggtttgaaa tcagaaagtg 22320
tgagtctttt ttgttgttgt tgttctccct tttcaggatt gttatggcta ttcttagtcc 22380
cttgtatttc catgtgaatt ttaggatcaa ctgtcaattt ctgcaaatga gccagctaga 22440
attttgatag agctacattt aatttttttt aatgaaccca cattgaaata ttcaacgatt 22500
gccaaaacat tatcacccca aacccactgt ttatcttaga gttcattctt ggtgttgtac 22560
atggatttgc catgtataat gacatgcatc caccattata gtatcatacg tgtagtttta 22620
ctgctctaaa aattttccgt gttccatctg ttcatccctc cccccaacac ctgacaacca 22680
ctgatggttc tattgtctcc atagttttgc catttgcaga atgttgcata tttggaatca 22740
tacagtatgt agccttttca gattggcttc tttcatgtag taatatgtat ttaaagttcc 22800
tccatgtttt tcatggcttg atagctcatt tatttttttag tgctgaataa cattccattt 22860
gcattgatat actgtagttt atctatttac ctatcgaagg acaccatagt tacttccaca 22920
tttcgatagt catgaataaa gctcctgtaa acatttgttt ctgggttttg gttggatgta 22980
tttttcaact cctttgggta aataccaaag ggtgcaattg ctggatcaaa tggtaataat 23040
atgtttagtt gtgtgagaat tatcaaactg tcaccctaag cataccattt ttgttgttgt 23100
tgttgttgtt attgtttttt gagatggagt ctcgctgtca cccaggctgg agtacagtat 23160
ggcgtgatct cagctcacca caacctccac ctccctggct caagcgattc tcgtgcctta 23220
gcctcccgag tagcttggat tacaggcacg tgccaccatg cccagctaat ttttgcattt 23280
ttagtagaga cggggtttca ccatgttggc caggctggtt tcaaactcct gacctcaggt 23340
gatcctcccg ccttggcttt ccaaagtgct gggattactg gcgtgagcca ccgtgcctgg 23400
tcgcaaagca taccattttg cattcccacc agcagtaaat gagagtttct gttaccacac 23460
atccttgcca ggacttaatg tcagtgtttt ggattttcac cattctgata ggtgtgtgtg 23520
gtatttttg ttttaatttg cagttctcag tgcttctcag tacctcccac ccaccttatc 23580
tgggacagga tggctggagg ggctggagtt gggtagttct ctttccccag gtaggttaag 23640
ctctggtaaa agtttctccc agaactttat attaaattaa aagtttcgca tgaggataag 23700
acttattagg agcagaatgc tctggcatat ttaaaatggt ttatttttcct ctccgcctcc 23760
tggaagcact cacgtgagga cctggtagag gttttggagg taaaattcac aaaagtgtgg 23820
tggccccaat tgggtccact tggagttttt attttaattt tctttatttt ttatttacat 23880
tttttgtttg ttttttttgtt ttggagatag tcttgctctg cggaccagaa tggaatacaa 23940
tggcacaatc tcagctcact gcaacctctg ccccctccga ttcaagctat tcttgtgcct 24000

cagcctctct cgagcagctg ggattacaga tgtacaccac catgcccgac taatttttgt 24060
atttttgta gagatggggt ttcgccacat tgcccaggct ggtctcaaga tcctagcctc 24120
aagtgatccg cctgcctcag cctcccaaag tgctgggatt acaggcgtga gctactgtgc 24180
ccagccttga ctttttaata atagccattc tgactgatga gatggtatct cattgtggtt 24240
tttgatttgc atttctcttc ttatttcttt tcttttcttt tctttttttt ttttttttt 24300
gagacaaagt ctgttgctga tgctcaggct ggagtacagt ggtgcaatct tggctcactg 24360
aaacctccgc ctcccagggt caagcagttc tcctgcctca gcctcccgag tagctgggat 24420
tataggcacc caccaccatg cccaactaat ttttgcattt gtagtagaga tggagtttca 24480
ccatgttgcc caggctggcc tcgaactcct gacctcaggt gatccgcctg ccttggcctc 24540
ctaaagtgct gggattacgg gtgtgagcca ctgcacccag ccaatttgca tttctctagt 24600
gatcagtgat aatcggaatg gcacgcaatt tacaacatta gttgtttatt tctggaattt 24660
tccatatagt attttcagac catggttgac cacagttaac tgaaaccttg caataagtgg 24720
ggactgctgg agatcctctt cattttaaag atgcagaaat aggttgggca gtagagcagg 24780
gcaactagga gaaatgcatt taaacaatag ggataactgc acagtgaggc agaaaatgtt 24840
atataataca ctagatttgg ctcaaaggca gtcacaggag cctctaggag ttcatttcca 24900
gtagagtcca ggagaggaat gagatactgg aattaagtag gttatacagt gatacttgaa 24960
gtaatgagac ttctttaaaa tgaaatgggt cttctaaaaa tgtttcataa atcctaaagg 25020
taggcacaac atatggcaca taaattatca gaaataccca tatgtataac atctcttttt 25080
tcatttggaa gtaaacagcc ctgcctcagc aaatttctcg agttaataaa atgtttttag 25140
tgtccccgtg tcttcttttc tattctcagt tggtgatttc ctttcttttg ccaatctccc 25200
tgataggcaa tctgtgctct ctgtctaacg tttctctctt gaatttatag tgtagttatt 25260
actaaggcac caattcctat caattaaggc tcatttcagt gtaatgtgga tgtgttttat 25320
ttagtattat gtccactggg cctttgtata tttagcgttt gcaccactcc ccaccccatt 25380
ttcatgtgac aactggaggc cttgttcaaa aaactaccag aaaaaggaca tgtatcccat 25440
cctttatggc tttcttgtaa caccagaaac tatccagcct tgttttgtcc tgtttactgt 25500
aagatcctaa gccagtggtt ctcaaccagg atgaatttgc tctccaggaa acgtttagca 25560
aagtctagag acatttttttg gttgtcacaa tgtggcaggg aagagggtgc cactggcatc 25620
tagtggatgg agagaactgc caaacaccct gtgatgtaca ggacagcccc tacaacaaag 25680
aattgtctta ctaaaatgtc agtattgccc aggttgagaa gcccttcact aagcaagccc 25740
cttttcctct ctaacctcag ttaggtcctt aattaagttt gtttcttcaa ctgaaaagtg 25800
gaggttattg caacttttttc aaatggttgc aaagataaat gaatttaagg cgtggaaata 25860
tttagcacag tatctgacac acagtggcac tctgctgaaa gaataaatta gggtttttaa 25920
aaaatgttgt ataatgcagc ttctcatcca tcctgatttt cctgcaaaaa ttctcatgtt 25980
aaatgttttg atccattatc actgtaaatc tcttacatgt tatagaaatt cctgtattta 26040
agcttctaaa atttcctaga ctacctctta aatatgtttt gatttttggt catgaaagca 26100
atgattatac caccactgga acaaatcttt ttaagtaaca gaaacctttt cctgttcttt 26160
tctaaagcac ttcagtgaat taaagattgg tgaggagtgt aggtcaaagt ttaagtcaat 26220
cagtaattga atattttctg ttgtcaaaga gttaataagt tcatgatcaa atagcgtatg 26280
taggagaaga cgttagaaag atttcagaaa cactaaactt atcgcattct ctttgtgctt 26340
ttaattttat gtagcaattc aaggccatca atagttgtag tcttaaaagc ccctaaaagc 26400
agaaatgata ttactttgtg gattgtttag atttttatga tgtacccctg tttgcaaggt 26460
ggtgttaggt cctatgaatt aggagtttgt aagtttagtt aattggctga ggcccattgt 26520
tgatctgtta attttttttgt gaattagttt ttaacatacg gatcctgcac aagataaatt 26580
tataccttta aacatttcag ttttttggta tcgtaaatgg tgttcaaact ttttattttg 26640
gcttccagtt cattgctggt atttacaaat acaattaatt tatgtgtgtt cttttaacct 26700
gcaaccttgc tgaattcatt ttaggaggtt ttggttagag tccttgggat tttctaaata 26760
ggctcatgtc tgtttgtctc tctttaatta aaatttttta gggtgtgttt ttgtaaagcg 26820
ttaatattca gttttgggtg ttttcattct cagactgatt ctcaacctgg ccctctttca 26880
aactttcaca ccttctccct tttcaacttc actcccttcc ccagaaatca ctgaaatgac 26940
ttctgagcac attgtcattt tatcttactg attcaaaatt attattagat acgttattat 27000
accctggagg taacagaata gtgcatccag agtctttaat tatgtaggac aattgtgtta 27060
gtgggtgaaa gctgttaggc ttttggattc ttgcacctac tttttattga ttttttttc 27120
ctcattgtca gttcttcatc tgctgttagt gtgtctgccc ttaaactacc tcttcctact 27180
tctaataatt tggcatgttg gaaaaccaga tactcaagct catcacagta ttttctagca 27240
actcccaata gagagatccc agaggcagtg gcagttaagt tgcattttca cttgtgaata 27300
aactattctg gggtttagat acataggtac tttatatttt tttaagttaa aaatttttta 27360
aagtttggaa caaatttgct taataaagtc agttttagcg ggaaattcac atgcacatga 27420
gaatcgaaag aaagaacctt taatgttgat gtaagcttat cattcctttt tcacatatac 27480
ctcaccacct cctctgtgtt taattatctg gacttttttc tcatccctag attttgtttc 27540
tcaaaaatgt ttctccctta aaacatgttt taatgaactc tgtaagatct aagttcttaa 27600
tctcaaagta atctttagat tgaggagtac taaaattgag atccacagac cctttaaggt 27660
cagggataga actgctttca gtgtgattga tttccttagt aatcttatgt gttttattca 27720

aaaacgttat ctgaaatagc tttcattgga ctgcataagg tttccatggc aaaaaaaaaa 27780
agtgttaaga atccgtgaga ggcaatccca gagtgagctc ctgtcattgc tgccagaaac 27840
cacaaagcct tttgagagtt ctgaactctg agtcacagag ttctgatagg tatgtggtat 27900
tgtaattcag ttgaattttg ctgttggaat atttaggtaa attgtcagac gtaattaaga 27960
tttgcagcat gaaatgggcc caaatatata tggagtcttg ctctgttgcc caggatggag 28020
tgcagtggca cgatctcagc tcactgcaac ctccacctcc caggttcaag tgattctcct 28080
gcctcagcct cccaccacgc cctgtctgtg ataagatttt ttttttttt tcttgagacg 28140
gagtcttgct ctgttgccca ggctggagtg cagtggcacg atctcagctt actgcaagct 28200
cgtcctcccg ggttcacacc attctcctgc ctcagcctcc ggagtagctg ggactacagg 28260
cacctgccac cacgctcagc ctgtgataag atttttaaat aaaaatttca cggctcgttg 28320
ctatggctca tgcctgtaat cccagcagtc tcggctcact gcaacctccg tctctccggt 28380
tcaagcgatt ctcctgcctc agactcccga gcagctggga ttacaggcac ataccaccac 28440
ccccagctaa ttttttctgt tttagtagag atggagtttc actatgttgg ccaggctggt 28500
ctcgaactcc tggtctcagg tgatccaccc gccttggcct ccctagtagc ctgaggctac 28560
aggcatgtac caccactcct ggctaatttt tgtagggatg gggtttcgcc atgttgccca 28620
ggctgggatt gcgggtgtga gccactgcac ttggccccaa aacaattttt aacggatgat 28680
gctgggacac ctacatataa ccacatagtt aagtgaagtt gggccccttt cttactatat 28740
gcaaaaatta actcagaatg gttcatacac ctaaatgtaa gagctaaaac tttataaacc 28800
tcttagaaga tataggagca aatgctcatg accttgggtt tgacaaactt cttaggatac 28860
tgaaagtaaa agtaaccaaa gttgaatttc attaaaattt aaaatttgtt cttcaaataa 28920
cgtaaagaaa gtgaaaagat aacccatagg atgcaagaaa atatttgggc ttttttttgt 28980
ttgtttgttt gtttgagaca gagtctccct ctgtcatcta gactggagtg cagcagcaca 29040
atcttggctt attgcaatgc cacctcccag gttcaagcag ttcttgtgcg tcagcctccc 29100
aagtagttgg gattacaggt gtgcactacc accgcctgct aatctttttt ttttttttgt 29160
ttttttttgt atttttagta gagacaaggt ttgaccatgt tggccaggct ggtcttgaac 29220
tcctgacctc aggtgatctg cctgcctcag cctctcgaag aaaatacttg taaatcattg 29280
gtctgctaag ggacttgtgt ccaggagata acaaaaagat aatttgctat cagagagggc 29340
aaatcaaaat cacaatgaga tatcacatct tatccacttg aatggttata ataaaaagat 29400
taagtttaag caatgctgta gagaaactgg aacattctat ttccctttt aaaatgcaca 29460
cttcattgat ttccagtgtg ttcagagtta cacaaccatc acttaccaat acagctctcg 29520
atgggattgt aagatggtgc aggtgctttg gaaaagtttg gcagttctac aaaaggtttt 29580
gcagagttcc attcattggt ttatataccc aataaaatga aaacatgtac aaaaacttat 29640
acatgtattt tcataacact gttattcaca gtagctaaca agtagaaaca acccatgtcc 29700
atcagtgact aaacaaaatg gtggtggact attcagtgat aaaaaggaat gcaattctga 29760
taaatgctgc atcatgggtg aactttgaaa acatgctaag tgaaagtcga gccacaaaag 29820
accacatatt agtatgattc catacacagg cataccttgt tttattgcac ttcacagggt 29880
tgtgtgtgtg tgtgtgttta cacaacttga aggtttgtgg cagctctgca tgaaacaatt 29940
cttttggcac cattttccca gcagctcatt tcatttctct gctacatttg gtaattctcg 30000
tgctttcaga ctctttcatt atatctttta cagtgatctg tgatcagttt tattagggtg 30060
gtctggaacg aaacccacac atctgaataa tgcttatcta tgaaatgccc agaataggca 30120
aatctataga gacaaaaagc aaattagtta gtagtagcct agggccgggg agtttggggg 30180
agtgactgtt aatgagtaca ctttcttttg ggggtaatga aaatgttcta aaattgataa 30240
gtgatggttg catagctctg aacacactgg aaaccattga agtgtgtatt ttaaataggg 30300
aaatagaatg tgaattatat ctctaaggtg ttacagttaa tggaagcctt tcaccttatc 30360
gtgatactgt atttcctcag tgctgagatg tgcttttaa ggctaggctt ggctgctcac 30420
tcctgggatc ccagcacttt ggaaggccaa ggcaggagga ttgcttgagc ccaggaattc 30480
tagaccagcc tgggcaacat agtgaaaacc tgtctctaca aaatgtatat ttttaattag 30540
cggggcatag tagcatgggc ctgtagtccc agctactcag gagactgaag tgggaggatc 30600
acttgagccc ggggggtcgt caaggctgca gtgagccaag attgtgccac tgcactctag 30660
cctaggtaac agaatgagac cctatctcaa aaaaaaaaaa aaaaaaaaaa aaaaaaatat 30720
atatatatat atatatatat atatatacac acacacacac acacacacat ttatcaggaa 30780
gatagatgtg atgctagctt ttacatctct ctgtgtaatt gttcaggaac tgcatcagtg 30840
ctaatctggg ggtttttgtt ttggtcaaga tttcatggaa catcagaagt agaacataca 30900
ttgagaccaa ctattctaaa tcctccattt tatggtctta aatgtcagtc tggcaaaatt 30960
cagagaaaac atttcccagt aggagactgc agctgggaag aaggtggtgc cagcagcagc 31020
ttgtaaggat ataaagagaa ggatcagaag agctggctct tatttcctac tggaggaaaa 31080
gataccgctt ttggtgacaa gctgcatgct ttagcagtga ctttatttt ccttcaccac 31140
aaaccagact tgtgctattt ccagcctcag cacattattt aggcacttcc ccactgcatt 31200
gttttcgttt ctgaaattgt atcacattcc ctaaattgaa tccaactttt aatgtagtgc 31260
ccttcagaca gacacgtttt acattcacct gttttgtgtc tttcttggct taaaccacct 31320
aagattaggg cccgatacat tcttgacaac atccatcatt ttataggtgc ttcatttata 31380
tattaacaaa caaccggaat gcttcagaga tttttcgcat cagatcagac cttgctgttt 31440

ttgatgtttg accctatctt ctggccgggt gcagtggctc acacctgtat cacacctgta 31500
atcccaacac tttgggaggc caaggcaggt ggatcacttg aggccaggat tcgagaccag 31560
cctggccaac atggtgaaac cctgtctcta ctaaaaataa aaaaattagc tgggcgtggt 31620
ggcgcgtgcc tgtactccca gctactcgag gttgaggtgg gagaattgct tgaacctggg 31680
aggtggaggt tgtagtgagc caagatcgcg ccactgcatt ccagcctggg tgacagagtg 31740
agaccctgtc tccaaaaaaa aaaaaaaaaa gactgtatct tctgtcctgc ttcagcatat 31800
tttataaccc tgtcttccat ttctattcta ccccccttcca tgctgctgca tttttccaca 31860
atgcatattt taattttaaa ttattccttt tctccttttc ttaaaggcag atgatttcct 31920
ttggattctc gctctttctc tcttttgcat gcagacatat ggagaactaa agaaagtatt 31980
aaggaaaaga ttgtttaaat ttatttgaaa tttgcttctc tgtgaattcc atcttgaaat 32040
ttaaactcac agggcaagcc catgtgtctc ctaatcaaca acgaggaaat ggtagttagg 32100
taagactgta taagacattt aagtagatga tagtggaaaa gggaagagag gaagtcagag 32160
gaatcaggaa ggatgagatg tctatctcca gtgactagtg gtgtcattag agaagagccc 32220
atatattaag taagccctta tcttggtaag tgctttactt atctgcttat tccttacaat 32280
actttttttt ttttttttt aacagagtct agctctgtgg gccaggctgg agtgcagtgg 32340
cgtgatctca gctcactgca acctccgcct cccaggttca agcaattctc atgtcgcagc 32400
ctccagagta gctgggataa cagagctaaa tgactgcatc tggcctccaa ccctttatta 32460
agataggtat tttgtctgtt atgctgatga ggaaccttaa tgctcagaga tattaagaag 32520
cttgcgtaag gccacataca taattaaata ccagagcaga aattccaggg aatgcctgta 32580
ctctttccat tatgtatttt tgttcattaa cctaggtagg gggcatagaa attctgagtt 32640
tgagaggcat gtggaaggaa gtatatttgg tttggttttg gacctattga agttcctgtg 32700
ggatgggtag gtggagtttt gaaagagaca acaacccagg tattttcact tgggagaatg 32760
ttgggtgtaa tttatatcaa gttttccttt ttggttttat tgttatttaa tccaaatcat 32820
tagcagtttt tctttttttt ttctttttga gacagcagtt tttctttttc ttattgttat 32880
ttaatccaaa tcattagcag tttttctttt ttctttttct ttcactttgt cacccagtct 32940
ggagtgtagt ggtgtgatct cgtctcactg tagctcaacc tcctgggctc aagcgatcct 33000
gctgcctcag tccccgaggt agctgggact acaggtgtac atcaccactt ctggctaatt 33060
tttgtatctt ttatagagac ggggtttcac catgttgccc aggctggtct caaactcctg 33120
agctcaagct atccgcctgt cttggccttc caaactgcaa ggattacagg cgtgagccac 33180
tgtgccaggc gcattagcaa tttttcatag catatctaca aagttctaat atgtatataa 33240
tgtatattac aagaacataa gtgctttgat ttaggttatt cagttatata attatgttga 33300
gtcccaactg tgtgccagat accgtgctga gtaggtagta aacattggta agcaaaacag 33360
acatagcccc tgcccttatc agactaccca catagcccgg acacagtggc tctcacctgt 33420
aatctcagca ctttgggagg ccaaggctag aggatcgctt gaggccagga gtttggtcag 33480
cctggtcagc gagacactgt ttctaccaaa acaaaaaaat tgctgggcat ggtgttgtgt 33540
acctgtaatc ccagctactg gggaggctgc agtgggagga ttacttgagc ctagaagttg 33600
gaggttgcag tgagatgtgt ttgtatacct tttgtcatta ggtccatatt taggccctga 33660
aataggctaa tcaaaatatt tattttttt tcatttgttt tcagctgggg acgaaacaca 33720
gtagctttgc aaatataaag tgggagcaca acctagcact ggcgttggag gtcctcaaaa 33780
cagtccccag gtttggtgat tcaataggag ggctcttagg acttagtata taatcttact 33840
cacatctatg atttgttaca gcaaagaatc aaagcaaaat cagctaaggg gaaaggtgct 33900
ttggtctaaa gggaactaag cacaagcttc caagagttat ctcctagtgg agtcacacag 33960
gacatgctta attccttcaa catatatgaa aggttttctg ccagagaagc tcatttggag 34020
actcagtgcc ccaggtactg attaggcacc cactgcctag catatgacaa aattcctgat 34080
tctcagaggg gaagcaggtg ttgaacataa accatatagt gttcgcagat agggaatggt 34140
gggaaccacc cccaaatcca tatcttttt tgttgttaag atggagtctc tctctgttgc 34200
ccaggctaga gtacagtggc aagatcttgg ctcactgcaa cctctgcctc ccgggttaaa 34260
gcagttctct tgcctcagct tactgagtag ctgggattac aggcgtgcac caccacgccc 34320
agctaatttt agtattttta gtagagatgg tgttttacca tgttggccag gctggtctcg 34380
aactcctgac ctcaggtgat acgcctgcct cggcctttca aagttctggg attacacgca 34440
tgagctacca tgcccagcca tcacatcttt taaaattgtt gtttttatac ccctgtctaa 34500
tcttctactc actcagtgtc tgttagccaa caattgaatc aggattccat tttgctgcct 34560
agaaaatcag tggctgtcta gttatgaaac aattgataac tcagatatca agccatttta 34620
aaacaagtat catgggaatt aaaatttaaa attaggcatg tgtggtaaga agggtgaggc 34680
cacattcatc tttaaaatat caggaggtct tagtcatttc ctttcctcta tggttttctg 34740
aaccatgtaa actgaaatta gattgttcaa tcattttaaa tatctaaaag atatagagag 34800
ggttcctttg ttttaacaaa aaaataccaa aggagaaaat acactcataa cgtgaattaa 34860
aatatagttt ttatttttctt ctttaatgtt tattaatgtt tccccgtaaa gtagaacttt 34920
ccttttctgt tctgctttct gtgttctcta cacgtatgtg aaccacaaga ctgaagaaat 34980
ctagcctctg aaaaacccat agttggagtt tattttctct aaaacaattt agttatttttt 35040
gtttgttagg ttcaaggatt taagagaccc tttatattat cttagtctta aaccaatttg 35100
ttcccagttt gcgtgtggat gacagctcat caaaactggt caactaggac tgaggctaag 35160

attcaggaag gtaatttaga ctggaaatgg taattcaagg gttgaagaac tcagagctgg 35220
cagagatcac aacagagtcc aagagtttgt agtattcaaa gagaagccag aaagccagga 35280
ccaaagccaa ggacagtagc acttagctta aagtctgcaa agcagagaat catgatgggg 35340
agaatcatga cgttggggag atgttagaat aatttacaaa tattggcatt tacaaagctc 35400
ttagttatct catgattata tcaactcttt aaagtgggta ctgttatctg ctttattcaa 35460
agacattaag agacttggtc caagcaatca cagtaagtag ccgagctgag actgaaatct 35520
tataatttaa ttcttcagac tgaaaatctc atcttctgag atagtagtgt ttttcagact 35580
gctctgttga ttctgaggat tctgcaaagg taaccaaggt gagggcttgg ggatggggag 35640
cctcatccct tctaggccac actccatccc agcttcattc catcagagtt gctcagtttt 35700
aaatattatt ataggccagg caccatggct caaccctgtg gccccactac tcaggaggct 35760
gaggttggag gattgcttga gcccaggggg ttgagtttac agtaagttgt gtttgtgcca 35820
ctgcacttaa gcctgggcaa cagagggaga ccctatctca aaaaaaggtg ggggttatac 35880
acacgattcc tttgaattgt tttgttgtgt aaaaaatcag ttggattagg agcagcagaa 35940
ttctaatttc agtattgtca attagtgact gtaggcaagc ttggtcctat gaaatgggaa 36000
agttagtctt agataacttt tagatttatt tttgccatag tccagtccat ttaaaaagtt 36060
ttaaaaatta tttttagaga tgaggtcttg ctatgttgtc caggctggtc tcttatcatg 36120
tggaaatagt tcatttaaca ttccatgcca gttcagtgaa gtctgtgttt atatccttta 36180
ggcagaatga gaatgttttg gtgtgtgtat gttttattct agtagctgaa gttaaggcag 36240
ggtgaaacac agtcatacag ccaaacacta gcccagcttc ctacccctcc cagggctccc 36300
tttcccttct atctctgacc aaaaaataag ccctcgaaag gtttttaaaa aataatcttt 36360
ttcctctata ttttaatatt ttaggaacaa gttccagttc cagtctacca cccaacacct 36420
agccagactc ggctagcaac tcagctgact gaagaggaac aaattaggat agctcaaaga 36480
ataggtctta tacaacatct gcctaaagga gtttatgacc ctggaagaga tggatcagaa 36540
aaaaagatcc gggagtaagt tttaattaat ttgtacattt caaagtttta ttttcagatt 36600
gcttttgaaa aatttgcctt ttctcctgca ctttagaaag tcagacattt aatatattga 36660
aacacttcgc ttaatgcaag agaatgttta attttttttc atagctctat tactggctac 36720
tagcaagaga atgttatgta gtaattaaga attttcacag gccgagcaca gtggcttatg 36780
cctgtaatcc cagcactttg ggaggctgag gtcgggggat cacttgaggc caggagttcg 36840
agaccaacct ggccaacatg gtgaaacctc atctctactg aaaatacgaa aattagccag 36900
gcatggtggc acgcgcctgt aatcccaacc acttgggagg ctgaggtggg agaatcactt 36960
gaaactctgt ctcaaaaagt aaaataaagt aaaataagag aaagttagtg ttaatgaaga 37020
gagctttaac ttacagcaac ccagtaagta ggctgaagtt agtgtttctg tttttacagt 37080
tattaaaatg aaaattatct tactctagga ttactctaaa tagcacaagg aaacaaagtt 37140
gcccgaagtt agtgtttctg ttttacagtt actaaaatga aaattgtctt agtccaggat 37200
tacactaaat agtgcaagaa aacaaagttg atctaactga tgcttttcta cctaaaataa 37260
tcccattctc cttcaggtat ctgctataat atgtagctgt ttacaagcac agcagtagat 37320
ttgatgcata tgtaatccat ttctctttgc tctgttgatt gctcctaagt cctgcaaatt 37380
ctcaggggat tgaaataaga gactgtggac agatcaacat taggaaaaaa ttcagtttgc 37440
atgctttggc aataagatga caggcaaagc attctgatgt gttactcaga ttaacactgt 37500
ctttaatggg atttaaagag ttttctgaat aggacacttc tgaaaatgcc ttctgatttc 37560
tctccattgc tctcaacagg tgtgtgatct gtatgatgga ctttgtttat ggggacccaa 37620
ttcgatttct gccgtgcatg cacatctatc acctggactg tatagatgac tggttgatga 37680
gatccttcac gtgcccctcc tgcatggagc cagttgatgc agcactgctt tcatcctatg 37740
agactaattg agccagggtc tcttatctga cttcaagtga accaccattt tggtggtttt 37800
gatcttttgt cactgagccc aaagagccag ggattaggaa ttaagatcgt gcacaaaagt 37860
ttccttaaaa ttcctggatg gctgcagatg ttgggggaaa aagtacgtga tattttagaa 37920
acttagtggg aaaagtagga tggtattttt atgtaaagcc ttgacccaat gtttaaaaat 37980
ataattgtat ttagatcttg ttattgctcc agtacatagg aattgtgtaa agtgttaaca 38040
gcagctgtat ttgtttaaat tgtgtgtatt gaagattagg aaaaagatag tagttatttt 38100
tcctaaatga aataactttc ttctcttccc cttccccacc cgaattcttt tctgaagttg 38160
ctggcatttg ggtcaaggtt ttattaaaag ctacatttta taacactggc acacacaaaa 38220
aagtagtttt aagcttgttt gcacagttct ttttttccat tggaaatgga attcattgcc 38280
ttaggtcttt ttaaatagtg tattattatc gttggggctg gctctatgct tgaaaaccag 38340
tttatttata acctgttata agtgctatat tctgtttgca gttaggaaat gcagaattca 38400
aagtgatctc ctagcttgta agcaaactga gatgcactat ccctttttcta taaaaaataa 38460
gttaatgtgt caagaaacca actctattaa ggtggggttt aatattaccc tttcctatgt 38520
gttttatcta attattttgg ttgttaatat ggtgataatg gaaagtcaag ttaaatttta 38580
aatattaaga attctgattt attgagattg aattatgcca ccacgtttat gtaaaaatga 38640
aggtggcacc gtggtgagac ctaatgagaa atagttactc agttgtaaaa attttgattt 38700
attctctttc ttctgacctc cttgcctctt gtcttgaacc atagcaaaag gatactgcat 38760
ctctcattac tgtagtgctg aggttattga agttatacaa aacacatctc agtctctgtt 38820
tcttggaaag gtatctatta catcctgcta gctgactgac aaaactaagc agggagaata 38880

aagataattg tattttatgt tttgcacaca aacgcagaat ttgtataacc atatgacttc 38940
atagttgtga tctcaaaaaa gaaggaattt ctcctttgtt tcttgcagtt aatgtaagaa 39000
tactttaaat ctctaagctt ctgaagtgtt agaggtagag atggtctagt aaagatgtag 39060
tagtaatgtt ttatccattt agcatgtgtt tatttttca tatgtactca aaggtgactt 39120
attggttcac ctcagtgata ttacagctaa aaaaatcatt cattagcaaa aggaaaagtg 39180
gtctcaacct aacatcagaa gtgtttctta ttattatttt atattgagtt gaatattgaa 39240
ctctaacagt tttctacata caaaacacag tgtcatgaag gttattcata attgcattat 39300
agaggaatgt agtatgtcat aagtactttg taaagatttg acattcaact gtagtatcca 39360
tatgttgctt aaatttcctt atgagcccca tgatggaaag acttaaagat gaatttgaga 39420
aaaattgaaa gaaattagat tatcaggttc tgttaaattg ttacatgtat cttgcttaaa 39480
tttctgttta ttaatttata tccacccaag tacataaagc aaatttggag gaaacaactg 39540
aagttgtgca atattttctg ataattgctt tttttattct tgtgttttct acttaaacat 39600
aatgtctgtg tcatcaagta ttatagtcag acttttcttt ttttctagat tgttaaaatt 39660
ggcaaatgaa ctttttttaaa aatcatcttc catgttgcag ttagtctttc ttttcattac 39720
aagtctttca cagaagtttg gtggtaatat tgaaagaact agcattgggc agaatgtgtc 39780
ttttt 39785

<210> 2
<211> 154
<212> PRT
<213> Homo sapiens

<400> 2
Met Gly Asn Cys Leu Lys Ser Pro Thr Ser Asp Asp Ile Ser Leu Leu
1 5 10 15
His Glu Ser Gln Ser Asp Arg Ala Ser Phe Gly Glu Gly Thr Glu Pro
20 25 30
Asp Gln Glu Pro Pro Pro Pro Tyr Gln Glu Gln Val Pro Val Pro Val
35 40 45
Tyr His Pro Thr Pro Ser Gln Thr Arg Leu Ala Thr Gln Leu Thr Glu
50 55 60
Glu Glu Gln Ile Arg Ile Ala Gln Arg Ile Gly Leu Ile Gln His Leu
65 70 75 80
Pro Lys Gly Val Tyr Asp Pro Gly Arg Asp Gly Ser Glu Lys Lys Ile
85 90 95
Arg Glu Cys Val Ile Cys Met Met Asp Phe Val Tyr Gly Asp Pro Ile
100 105 110
Arg Phe Leu Pro Cys Met His Ile Tyr His Leu Asp Cys Ile Asp Asp
115 120 125
Trp Leu Met Arg Ser Phe Thr Cys Pro Ser Cys Met Glu Pro Val Asp
130 135 140
Ala Ala Leu Leu Ser Ser Tyr Glu Thr Asn
145 150